



CIÊNCIAS EMPRESARIAIS

ESCOLA SUPERIOR
POLITÉCNICO SETÚBAL

PAULO JORGE
ARAÚJO OLIVEIRA

AVALIAÇÃO GLOBAL DA ESPERANÇA MÉDIA DE VIDA E FATORES QUE CONTRIBUEM PARA O SEU AUMENTO



Relatório de Dissertação/projeto/estágio/projeto
de investigação do Mestrado em Ciência de
Dados para Empresas

ORIENTADOR

Professor, Hernâni Raul Vergueiro Monteiro
Cidade Mourão

Novembro de 2025



PRR
Plano de Recuperação
e Resiliência



REPÚBLICA
PORTUGUESA



Financiado pela
União Europeia
NextGenerationEU

PAULO JORGE
ARAÚJO OLIVEIRA

**AVALIAÇÃO GLOBAL DA
ESPERANÇA MÉDIA DE VIDA E
FATORES QUE CONTRIBUEM PARA
O SEU AUMENTO**

JÚRI

Presidente: Prof. Coordenador Ana de Jesus Pereira
Barreira Mendes, IPS

Orientador: Prof. Adjunto Hernâni Raul Vergueiro
Monteiro Cidade Mourão, IPS

Vogal: Prof. Adjunto Vítor Manuel Meneses Barbosa,
IPS

Novembro de 2025

Avaliação global da esperança média de vida e fatores que contribuem para o seu aumento

Paulo Oliveira, 2025

"É provável que no século XXI os humanos empreendam uma tentativa séria de alcançar a imortalidade. Lutar contra o envelhecimento e a morte será apenas a continuação da luta mais antiga contra a fome e as doenças evidenciando o valor supremo da cultura contemporânea: o valor da vida humana".

Harari, Y.N (2017)

Agradecimentos à minha família por ter sempre dado o suporte coletivo para a realização deste desafio pessoal.

RESUMO

Numa época em que se conseguiu reduzir os índices de pobreza, guerras e doenças, o tema da longevidade do ser humano está em destaque com o constante desenvolvimento ao nível da medicina, bioengenharia e sistemas de monitorização dos indicadores de saúde.

Foi com base nesta temática que surgiu a ideia de perceber como tem sido a evolução da esperança média de vida do ser humano a nível mundial com o objetivo de perceber quais os fatores que a influenciam.

Nesse sentido irá recorrer-se aos dados disponíveis nas seguintes plataformas:

<https://ourworldindata.org/>

<https://databank.worldbank.org/>

<https://www.pordata.pt/pt>

<https://www.mortality.org/>

Com base na análise dos dados mais relevantes para este trabalho será feito o respetivo tratamento e preparação para uma posterior análise a partir dos modelos de aprendizagem supervisionada.

Irá ser feita uma primeira análise para perceber como a esperança média de vida tem evoluído até aos dias de hoje, como se encontra atualmente e qual a tendência para o futuro.

Também se pretende criar um modelo de aprendizagem supervisionada que nos permita prever a Esperança média de vida com base em vários indicadores independentes.

Para verificar como estão os países com o valor mais alto da Esperança média de vida em 2023 em cada uma das categorias analisadas, serão criados gráficos que permitam tirar conclusões sobre quais os indicadores comuns nesses países.

Complementarmente far-se-á uma análise comparativa entre os dados relativos ao nosso país por áreas geográficas.

Procurar-se-á também expor os resultados de uma forma gráfica simples, mas clara para uma interpretação visual mais eficaz.

O projeto será desenvolvido de forma a aplicar o máximo dos conhecimentos adquiridos durante o 1º ano do Mestrado de Ciência de Dados para Empresas por forma a desenvolver e aplicar esses conhecimentos da melhor forma.

No final será feita uma análise a todos os resultados obtidos e serão tiradas as conclusões possíveis com esses mesmos resultados.

PALAVRAS-CHAVE

Esperança média de vida; Aprendizagem Supervisionada; Saúde e bem-estar;
Geo-analítica

ABSTRACT

At a time when poverty, wars, and diseases have been significantly reduced, the topic of human longevity is in the spotlight, thanks to continual developments in medicine, bioengineering, and systems for monitoring health indicators.

It was against this backdrop that the idea emerged to understand how human life expectancy has evolved globally, with the aim of identifying the factors that influence it.

To this end, data available from the following platforms will be used:

<https://ourworldindata.org/>

<https://databank.worldbank.org/>

<https://www.pordata.pt/pt>

<https://www.mortality.org/>

Based on the analysis of the most relevant data for this project, the data will be processed and prepared for subsequent analysis using supervised learning models.

An initial analysis will be conducted to understand how life expectancy has evolved up to the present day, its current state, and the trend for the future.

The project also aims to develop a supervised learning model that allows us to predict average life expectancy based on various independent indicators.

To examine the countries with the highest life expectancy in 2023 across all analysed categories, graphs will be created to help identify common indicators among these countries.

Additionally, a comparative analysis will be carried out between the data related to Portugal, broken down by geographical areas.

Efforts will also be made to present the results in a simple yet clear graphical format for more effective visual interpretation.

The project will be developed to apply as much of the knowledge acquired during the first year of the Master in Data Science for Business as possible, in order to develop and apply these skills in the best possible way.

Finally, an analysis of all the results obtained will be conducted, and conclusions will be drawn accordingly.

KEYWORDS

Life Expectancy; Supervised Machine Learning; Health and Well-being; Geo-analytics

ÍNDICE

RESUMO	V
ABSTRACT	VII
ÍNDICE DE FIGURAS E QUADROS.....	XII
LISTA DE ABREVIATURAS	XIV
1. INTRODUÇÃO	1
1.1 Enquadramento ao tema.....	1
1.2 Formulação do problema.....	1
1.3 Definição dos objetivos	2
1.4 Abordagem metodológica	3
1.5 Resultados esperados	4
1.6 Estrutura do relatório.....	5
2. REVISÃO DA LITERATURA.....	6
3. OBJETIVOS E METODOLOGIA.....	10
4. ANÁLISE DE RESULTADOS	12
4.1 Esperança média de vida nos últimos 73 anos.....	12
4.2 Visualização dos dados	17
5. FATORES INFLUENCIADORES DA ESPERANÇA MÉDIA DE VIDA ...	22
5.1 Correlação individual dos fatores	23
5.2 Modelo de regressão linear múltipla.....	25
5.3 Modelo de random forest.....	30
5.4 Modelo de regressão linear múltipla II.....	33
5.5 Modelo de random forest II.....	35
5.6 Análise dos indicadores dos países Top 10 em 2023	37
6. PREVISÃO DA EMV PARA OS PRÓXIMOS 50 ANOS	40
7. ANÁLISE REGIONAL DA ESPERANÇA MÉDIA DE VIDA EM PORTUGAL.....	42
7.1 Preparação.....	42
7.2 Mapas temáticos	42
7.3 Dashboard interativo	47

CONCLUSÃO E INVESTIGAÇÃO FUTURA	49
REFERÊNCIAS BIBLIOGRÁFICAS	51
APÊNDICES	53

ÍNDICE DE FIGURAS E QUADROS

Figura 4.1 – evolução da EMV por continente	12
Figura 4.2 – EMV em 2023 por nível de rendimento dos países.....	13
Quadro 4.1 – Top 10 dos países com maior EMV em 2023	13
Quadro 4.2 – Top 10 dos países com menor EMV em 2023	14
Quadro 4.3 – Top 10 dos países com maior EMV em 2023 na Europa	14
Quadro 4.4 – Top 10 dos países com maior EMV em 2023 na Ásia.....	15
Quadro 4.5 – Top 10 dos países com maior EMV em 2023 na Oceânia	15
Quadro 4.6 – Top 10 dos países com maior EMV em 2023 em África	16
Quadro 4.7 – Top 10 dos países com maior EMV em 2023 na América do Norte	16
Quadro 4.8 – Top 10 dos países com maior EMV em 2023 na América do Sul	17
Figura 4.3 – Esperança média de vida global em 1950, 1980 e 2023	17
Figura 4.4 – Boxplot dos países com a mediana mais elevada	18
Figura 4.5 – Top 10 dos países com maior amplitude de valores entre 1950 e 2023.....	19
Figura 4.6– Top 10 dos países com menor amplitude de valores entre 1950 e 2023	20
Figura 4.7 – Mapa mundo da EMV em 1950 e 2023	21
Quadro 5.1 – Regressão linear simples de indicadores vários com EMV.....	24
Figura 5.1 – mapa de multicolinearidade dos indicadores em estudo	26
Quadro 5.2 – Regressão linear múltipla de indicadores vários com EMV	27
Quadro 5.3 – Valores de VIF dos indicadores no modelo	27
Quadro 5.4 – Regressão linear múltipla excluindo valores indicados acima	28
Quadro 5.5 – Previsão do modelo e verificação de exatidão	28
Figura 5.2 – visualização gráfica do teste efetuado.....	29
Figura 5.3 – visualização gráfica do método exaustivo.....	29
Figura 5.4 – visualização gráfica do método alternativo	30
Figura 5.5 – importância das variáveis no modelo Random Forest	31
Figura 5.6 – dados introduzidos para testar o modelo e verificar output	32
Quadro 5.6 – Teste executado para os dois modelos	32
Figura 5.7 – mapa de multicolinearidade dos indicadores do segundo estudo	33
Quadro 5.7 – Regressão linear múltipla do segundo modelo com todas as variáveis	34

Quadro 5.8 – Valores de VIF dos indicadores no modelo II.....	34
Figura 5.8 – visualização gráfica do teste efetuado com o modelo II.....	35
Quadro 5.8 – Regressão linear múltipla do modelo II ajustado com método alternativo	35
Figura 5.9 – importância das variáveis no Random Forest com o modelo II	36
Quadro 5.9 – Comparativo dos modelos.....	37
Figura 5.10 – esperança média de vida vs. Serviços de água potável.....	38
Figura 5.11 – esperança média de vida vs. Índice de desenvolvimento humano	38
Figura 5.12 – esperança média de vida vs. PIB agrícola	39
Figura 5.13 – esperança média de vida vs. Consumo de álcool.....	39
Figura 6.1 – EMV prevista para Portugal até 2073.....	40
Figura 6.2 – EMV prevista para a Suíça até 2073	40
Figura 6.3 – EMV prevista para Japão até 2073	41
Figura 6.4 – EMV prevista para Japão, Portugal e Suíça até 2073, segundo método logístico	41
Figura 6.1 – esperança média de vida à nascença 2021-2023	43
Figura 6.2 – esperança média de vida vs taxa de natalidade	44
Figura 6.3 – esperança média de vida vs taxa de mortalidade	44
Figura 6.4 – esperança média de vida vs densidade populacional.....	45
Figura 6.5 – esperança média de vida vs índice de sustentabilidade potencial	45
Figura 6.6 – esperança média de vida vs distribuição do rendimento bruto	46
Figura 6.7 – scatter plot matrix para as variáveis em estudo.....	47
Figura 6.8 – scatter plot entre EMV e taxa de mortalidade.....	47
Figura 6.9 – Dashboard criado no ArcGis online	48

LISTA DE ABREVIATURAS

EMV – Esperança média de vida

OE – Objetivo específico

VIF – Fator de inflação de variância (Variance inflation factor)

1. INTRODUÇÃO

1.1 Enquadramento ao tema

A esperança média de vida é uma medida demográfica crucial que reflete a saúde e o bem-estar de uma população. Esta medida, que representa a média de anos que uma pessoa pode esperar viver a partir do nascimento, tem evoluído significativamente ao longo das últimas décadas devido a avanços na medicina, melhorias nas condições de vida e políticas de saúde pública. No entanto, essa evolução não foi uniforme em todo o mundo. As diferenças entre países e regiões são marcantes, refletindo disparidades em fatores socioeconómicos, acesso a cuidados de saúde, educação, saneamento básico e outros fatores.

Historicamente, os países desenvolvidos experimentaram aumentos substanciais na esperança média de vida desde o início do século XX, graças a inovações médicas e melhoria nas infraestruturas de saúde. Em contraste, muitos países em desenvolvimento ainda enfrentam desafios significativos, incluindo doenças infecciosas, mortalidade infantil elevada e acesso limitado a serviços de saúde, o que tem impedido um aumento equivalente na esperança média de vida.

O estudo da evolução da esperança média de vida e das diferenças entre países é fundamental para entender as dinâmicas globais de saúde e para a formulação de políticas eficazes. Analisar esses dados permite identificar tendências, avaliar o impacto das intervenções de saúde pública e determinar áreas que necessitam de maior atenção e recursos. Além disso, compreender as desigualdades na esperança média de vida entre diferentes populações pode ajudar a direcionar esforços para reduzir disparidades e promover a equidade na saúde.

Neste contexto, o uso de técnicas de ciência de dados oferece uma oportunidade única para explorar grandes volumes de dados demográficos e de saúde de forma detalhada e sistemática. Ferramentas de análise estatística, aprendizagem supervisionada e apresentação de dados permitem não só descrever a evolução histórica da esperança média de vida, mas também identificar padrões ocultos, prever futuras tendências e fornecer “insights” acionáveis para tomadas de decisão.

1.2 Formulação do problema

Este projeto visa realizar uma análise abrangente da evolução da esperança média de vida ao longo do tempo, com um foco particular nas diferenças observadas entre países.

Utilizando dados históricos e contemporâneos, o estudo pretende identificar os principais fatores que influenciaram as mudanças na esperança média de vida e explorar as desigualdades persistentes entre diferentes regiões do mundo.

Através desta análise, espera-se contribuir para um entendimento mais profundo das condições que promovem a longevidade e daquelas que precisam ser melhoradas para alcançar uma maior equidade em saúde global.

Este estudo pretende responder a uma série de questões relacionadas com a esperança média de vida e apontar fatores que justifiquem esse aumento, tanto globalmente como a nível individual, nos países que se destacam relativamente a outros.

- O que contribui para que a esperança de média de vida tenha subido mais do dobro em cerca de 150 anos?
- Quais as diferenças entre os países que têm maior esperança de vida e os que têm menos.
- Há fatores comuns que justifiquem as diferenças?
- Qual é a expectativa da esperança média de vida para as próximas décadas?
- Há diferenças identificadas para as diferentes regiões de Portugal? Se sim, é possível identificar a que se devem essas diferenças?

1.3 Definição dos objetivos

Até 1870, a esperança média de vida situava-se entre os 29.7 anos no mundo e os 36.2 anos na Europa, de acordo com os dados disponíveis.

Esses valores aumentaram substancialmente e em 1950 os valores apontam para 46.4 anos no mundo e 62 na Europa.

A partir daí tem havido um aumento progressivo e em 2021 os dados apontam para os 77 anos na Europa, 70,9 anos no Mundo e por exemplo, 78,5 na Oceânia.

Olhando especificamente para determinados países, em 2021 a Esperança média de vida em Portugal registava 81,4 anos e no Japão esse valor situava-se nos 84.5.

O que leva a que estes dois países tenham valores acima dos registados globalmente?

Será possível relacionar a esperança média de vida com outros dados conhecidos e perceber quais os fatores que contribuem para o aumento da longevidade?

Haverá outros países que tenham valores acima da média? Que fatores contribuíram para que a esperança média de vida tenha aumentado e de que forma evoluíram globalmente?

Quais os países onde a esperança média de vida é mais alta? É possível encontrar indicadores que expliquem o motivo por estes países se destacarem entre os demais?

O objetivo geral do estudo é fazer as devidas comparações temporais e regionais e perceber de que forma a esperança média de vida tem evoluído e quais os fatores que para tal contribuíram.

Pretende-se fazer um estudo demográfico exaustivo, tendo em conta a informação disponível, e se possível, fazer uma análise a nível nacional para perceber se existem grandes diferenças entre regiões.

1.4 Abordagem metodológica

As principais fontes dos dados a usar serão os dados alojados nos sites ourworldindata.org, worldbank.org, mortality.org ao nível global e o site pordata.pt ao nível nacional.

Uma vez que se trata de um trabalho individual, e tendo em conta o estudo de Martinez, I. et al (2020) sobre as metodologias de um projeto de Ciência de Dados, a vertente de Gestão de Equipas não se coloca.

Ficam por avaliar as vertentes de gestão de projetos e a vertente de gestão de dados e informação.

De acordo com o artigo analisado sobre as metodologias de Ciência de dados, e descurando a componente de Gestão de Equipas, há 4 metodologias que apresentam pontuações no estudo mais elevadas:

4.10 Agile Delivery Framework (28)

4.19 Toward Data mining Engineering (27)

4.9 Big Data Management Canvas (26)

4.4 RAMSYS (26)

Após analisar as metodologias em causa, optou-se por escolher para este projeto a metodologia 4.9, já que prioriza a criação de valor a partir dos dados em vez de focar no armazenamento, poder computacional ou analítica.

Esta metodologia tem como fator negativo o facto de não endereçar a gestão de equipas, mas neste caso essa preocupação não se coloca, já que não se trata de um trabalho de equipa.

1.5 Resultados esperados

O objetivo principal é aplicar os conhecimentos adquiridos no primeiro ano do Mestrado e fazer uso das metodologias e ferramentas disponíveis no âmbito da Ciência de Dados para recolher, tratar, analisar, preparar e apresentar os dados no âmbito deste projeto.

Todas as Unidades curriculares contribuíram para o enriquecimento dos conhecimentos do aluno e a expectativa é pôr em prática todos esses conhecimentos, mas em particular os que estão relacionados com as seguintes Unidades curriculares:

- 1) Analítica de dados
- 2) Aprendizagem supervisionada e não supervisionada
- 3) Programação
- 4) Ética e privacidade de dados
- 5) Apresentação dos dados
- 6) Geo-analítica nas Empresas

Espera-se conseguir dar resposta às questões formuladas e conseguir obter informações valiosas que possam ser divulgadas e aplicadas pela comunidade em geral.

Acima de tudo pretende-se dividir o trabalho em 4 vertentes e obter resposta para as mesmas:

- 1) Passado – como evoluiu a esperança média de vida e a que se deve essa evolução:
- 2) Futuro – qual a previsão de evolução para as próximas décadas?
- 3) Como estamos atualmente a nível global? Existem fatores que se possam isolar e relacionar com as diferenças entre as várias regiões e países mundiais?
- 4) Como estamos atualmente a nível nacional? Há diferenças de região para região e conseguimos identificar fatores para essas diferenças?

1.6 Estrutura do relatório

Capítulo 2 – Revisão da literatura

Onde se abordam os principais artigos sobre a Esperança média de vida

Capítulo 3 – Objetivos e metodologia

Onde se faz uma explicação do que se pretende alcançar e de que forma

Capítulo 4 – Apresentação dos resultados

Onde se faz uma análise exploratória com base nos dados existentes no dataset da Esperança média de vida

Capítulo 5 – Fatores influenciadores da Esperança média de vida

Onde se criam os modelos de aprendizagem supervisionada que relacionam a Esperança média de vida com os fatores que a podem influenciar

Capítulo 6 – Previsão da EMV para os próximos 50 anos

Onde se procura usar outros modelos de aprendizagem supervisionada para calcular a Esperança média de vida nos próximos 50 anos

Capítulo 7 – Análise regional da Esperança média de vida em Portugal

Onde se procura visualizar as diferenças existentes nas regiões de Portugal

Conclusão e Investigação futura

2. REVISÃO DA LITERATURA

Foi realizada uma revisão da literatura vasta com recurso à consulta de artigos relacionados com a temática da esperança de vida e quais os fatores que contribuem para o aumento da mesma nas últimas décadas.

Este aumento tem-se registado em praticamente todo o mundo e é possível identificar algumas tendências globais e também disparidades a nível regional.

Para elaborar a revisão da literatura, foi criada uma matriz, que pode ser consultada no Apêndice 3, com os principais artigos consultados. Esta matriz identifica o tópico que abordam, por forma a estruturar a revisão por tópicos e fatores significativos.

Os campos principais que são enumerados para explicar o aumento da Esperança Média de Vida (mais à frente designado por EMV) são:

- Fatores sociais
- Fatores económicos
- Fatores regionais
- Saúde

Apesar de ser consensual que a EMV tem aumentado, não quer dizer que esse aumento corresponda a um envelhecimento saudável.

(Gulland, 2016) refere que o aumento da EMV corresponde também a um aumento de patologias do foro músculo-esquelético, ansiedade e depressões, complicações neurológicas e cardiovasculares e perda de visão e audição.

No mesmo sentido, (Crimmins, 2015) introduz o conceito de “healthspan” para se referir ao período médio de uma vida saudável.

Após reduzir significativamente as mortes por doenças infecciosas e mortes prematuras, o foco da medicina passou a estar concentrado nas doenças cardiovasculares e nos cancros na segunda metade do século passado.

Agora a população idosa enfrenta outros desafios relacionados com a saúde, incluindo doenças crónicas, enfarte do miocárdio e outras condições de saúde que devem ser equacionadas para que o aumento da esperança média de vida reflita também um prolongar da qualidade de vida.

Uma das questões levantadas em vários artigos científicos é se este aumento se irá manter ou se será atingido um limite a partir do qual não se prevê aumentos significativos.

(Le Bourg, 2012) formula duas hipóteses no seu trabalho para responder a esta questão. A primeira hipótese em que não existe um limite e a segunda, que esse limite existe.

No final acaba por refutar a ideia de que não existe um limite, uma vez que esse limite existe para todos os mamíferos.

Por outro lado, (Oeppen & Vaupel, 2002) defendem a ideia que a EMV pode continuar a aumentar no futuro, desafiando a noção de um limite máximo.

Para tal, destacam o facto de a EMV ter vindo a aumentar quase 3 meses por ano ao longo de 160 anos.

(Vallin J. e Mesle F., 2010) observaram que o ritmo de progresso da EMV variou ao longo do tempo em função dos avanços na saúde e da mudança de outros fatores que influenciam a longevidade.

Ressalvam ainda que embora o limite dos 100 anos seja possível, tal estará dependente das inovações futuras, que de momento são desconhecidas.

Fazendo um estudo sobre os fatores sociais que impactam a longevidade e na vertente dos fatores económicos e sociais, (Klenk et al., 2007) apontam que a reunificação da Alemanha Oriental trouxe aos seus habitantes rápidas melhorias na EMV, uma vez que as condições sociais ficaram equiparadas às condições da Alemanha Ocidental.

(Monsef & Mehrjardi, 2015) refere que a taxa de inflação e o desemprego tem um impacto negativo significativo e por outro lado os rendimentos afetam positivamente a EMV.

A urbanização é outro fator social apontado como influenciador positivo da EMV.

No mesmo sentido, (Yari et al., n.d.) explora os fatores que influenciam a EMV em diferentes países, destacando a importância de fatores socioeconómicos e políticos como o rendimento, educação, desigualdades e estabilidade política como contribuidores para o aumento da longevidade.

(Chetty et al., 2016) analisaram a relação entre os rendimentos e a EMV nos EUA entre 2001 e 2014.

Os resultados mostram que a EMV aumenta na população com rendimentos mais elevados.

Também são apontados fatores ambientais como a poluição do ar para o aumento da mortalidade.

Um estudo semelhante foi conduzido na China (Hao et al., 2020) com uma amostra de 27.794 idosos entre 2002 e 2014. O estudo indica que o acesso adequado aos cuidados de saúde está associado a uma EMV mais longa, com um aumento aproximado de 2 a 2,5 anos.

Ao nível regional há vários estudos interessantes que abordam perspectivas particulares como por exemplo no estudo de (Sugiura et al., 2010) onde se identificam os fatores que contribuíram positivamente para o aumento da EMV após a 2ª Grande Guerra no Japão : as reformas da terra, o trabalho comunitário, a ingestão de energia e proteínas, a educação e as reformas na saúde.

(Abbasi-Kangevari et al., 2023) analisam o impacto da poluição do ar na carga de doenças, mortalidade e EMV em 21 países da região do Norte da África e Médio Oriente e identificam que a poluição do ar está associada a uma série de doenças como doenças cardiovasculares, doenças respiratórias, diabetes, cancro do pulmão e outras.

Concluíram assim que a redução da exposição às partículas em suspensão poderia aumentar a EMV nas regiões estudadas.

Porém, os fatores mais evidenciados na literatura consultada são os fatores relacionados com a saúde da população, como seria de esperar.

Um conjunto de artigos (Mathers et al., 2015), (Buxbaum et al., 2020), (Monsef & Mehrjardi, 2015), (Kopeck et al., 2022) e (Lo et al., 2024) apontam os seguintes fatores como influenciadores da longevidade do ser humano.

- Tabagismo
- Doenças cardiovasculares
- Medicina e cuidados de saúde
- Dieta (em particular, consumo de frutas e vegetais) / Obesidade
- Exercício físico
- Redução de stress
- Acidentes de viação
- Pressão arterial
- Obesidade
- Colesterol

Além disso, (Reimers et al., 2012) levaram a cabo um estudo sobre a relação entre a atividade física e EMV. A revisão sugere que a atividade física regular está associada a uma EMV mais longa, com uma redução média na mortalidade de 31% a 35% em pessoas

fisicamente ativas em comparação com as inativas. Isso pode resultar num aumento de cerca de 3,9 a 4,4 anos na esperança de vida.

Adicionalmente (Fadnes et al., 2023) vão mais longe e sugerem que a mudança sustentada de padrões alimentares não saudáveis para padrões saudáveis pode aumentar a esperança de vida em 10 anos no Reino Unido.

3. OBJETIVOS E METODOLOGIA

Para definir os objetivos do projeto, a metodologia a seguir foi considerar a questão-problema de uma perspectiva mais abrangente e a mesma é apresentada de seguida:

“Quais os fatores que influenciam a esperança média de vida?”

Partindo da questão-problema estabelecida identificaram-se os seguintes objetivos específicos para o estudo:

OE1) Identificar como tem evoluído a EMV a nível mundial nos últimos 73 anos

OE2) Exploração dos dados multidimensional

OE3) Correlacionar EMV com outros fatores e procurar fatores comuns nos países com maior EMV

OE4) Identificar as regiões em Portugal onde a EMV é mais alta

OE5) Correlacionar E.M.V com outros fatores em Portugal

OE6) Identificar fatores comuns nos países do TOP 10 da EMV

OE7) Projetar aumento da EMV para os próximos 50 anos

OE1) Para realizar esta identificação recorreu-se ao site Our World In Data¹ e aos dados apresentados após se realizar uma busca por “life expectancy”. A fonte dos dados é UN WPP (2024), estudo das Nações Unidas apresentado no Prospeito da população mundial de 2024; HMD (2024), base de dados dos meios de saúde, também de 2024; Zijdeman et al. (2015); Riley (2005), atualizados a 3 de dezembro de 2024 e com um intervalo entre 1543 e 2023.

Os dados extraídos foram posteriormente transformados e carregados no PowerBI para se criarem visualizações gráficas que facilmente expliquem como tem sido a evolução da EMV nos últimos 73 anos.

Optou-se por este período, já que só a partir de 1951 é que as Nações Unidas publicaram com regularidade o WPP (World Population Prospects), relatório oficial com as estimativas e projeções da população global.

¹ www.ourworldindata.org

OE2) Visualização dos dados em vários tipos de gráficos através do Python para melhor entendimento dos dados, identificar lacunas e falhas nos dados e para orientação das próximas etapas.

Esta abordagem permite a limpeza dos dados encontrando dados errados, em falta, duplicados e outros.

OE3) Para a concretização deste objetivo serão efetuadas várias análises da aprendizagem supervisionada relacionando a esperança média de vida com outros fatores englobados nas seguintes categorias:

- Violência e Guerra
- Condições de vida, comunidade e bem-estar
- Inovação e mudanças tecnológicas
- Educação e conhecimento
- Pobreza e desenvolvimento económico
- Alimentação e agricultura
- Energia e meio ambiente
- Saúde
- População e alterações demográficas

OE4) Para este OE foram coletados os dados do INE relativamente à EMV nas diferentes regiões portuguesas e outros indicadores para permitir efetuar uma análise.

Por forma a conseguir visualizar corretamente estes valores recorreremos ao ArcGis (sistema de informação geográfica)

OE5) Ainda no ArcGis, foram criadas visualizações que nos permitem verificar os valores da EMV e outros indicadores, usando para tal, dois mapas distintos com cada um dos indicadores que são colocados lado a lado para uma mais fácil interpretação.

Também se desenvolveu um Dashboard interativo que permite ao utilizador visualizar e interagir com os dados regionais da EMV e taxa de natalidade em Portugal.

OE6) Para este objetivo pretende-se analisar a EMV de todos os países e procurar indicadores comuns nos países que têm a EMV mais alta.

OE7) Utilizar ferramentas de aprendizagem supervisionada para, com base nos dados históricos utilizados durante o projeto, estimar a EMV para os próximos 50 anos.

4. ANÁLISE DE RESULTADOS

4.1 Esperança média de vida nos últimos 73 anos

Analisando a Figura 4.1 que nos mostra a evolução da Esperança média de vida por Continente podemos tirar as seguintes conclusões:

- O continente norte-americano é o Continente onde a EMV é mais alta em 2023 e este registo manteve-se sempre desde que há registos em todos os continentes, ou seja, desde 1950, com exceção do período 2020-2021 associado à pandemia do Covid 21, onde foi ultrapassado pelo continente europeu.
- O aumento da EMV nos continentes asiático e africano foi mais acentuado do que nos restantes com percentagens de aumento na ordem dos 70%.
- Já os restantes continentes tiveram aumentos menos significativos (entre os 17% no continente norte-americano e 37% no sul-americano).
- No período entre 1959 e 1962 registou-se uma baixa muito significativa no valor da EMV no continente asiático.

Vários fatores contribuíram para este fenómeno, sendo o mais significativo a grande fome chinesa ou grande fome de Mao que se registou nessa época e onde 45 milhões de chineses passaram fome, foram sujeitos a trabalhos forçados ou espancados até à morte - Dikötter, F. (2012).

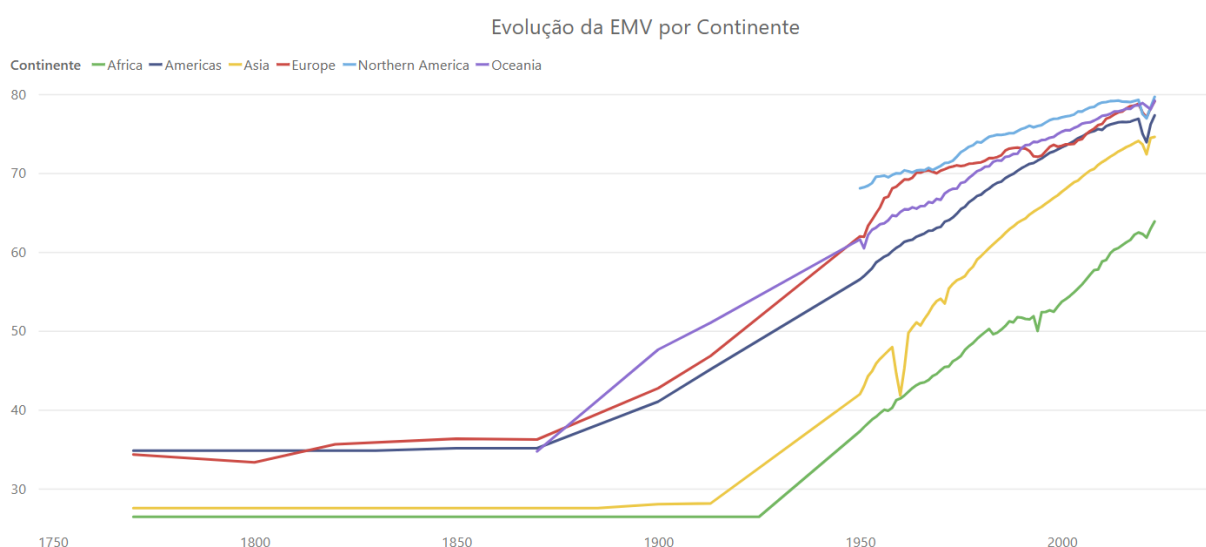


Figura 4.1 – evolução da EMV por continente

Na figura 4.2 podemos verificar que a EMV está diretamente relacionada com o nível de receitas per capita, já que observamos que o aumento do rendimento nos países é traduzido de forma direta no aumento da EMV nesses países.

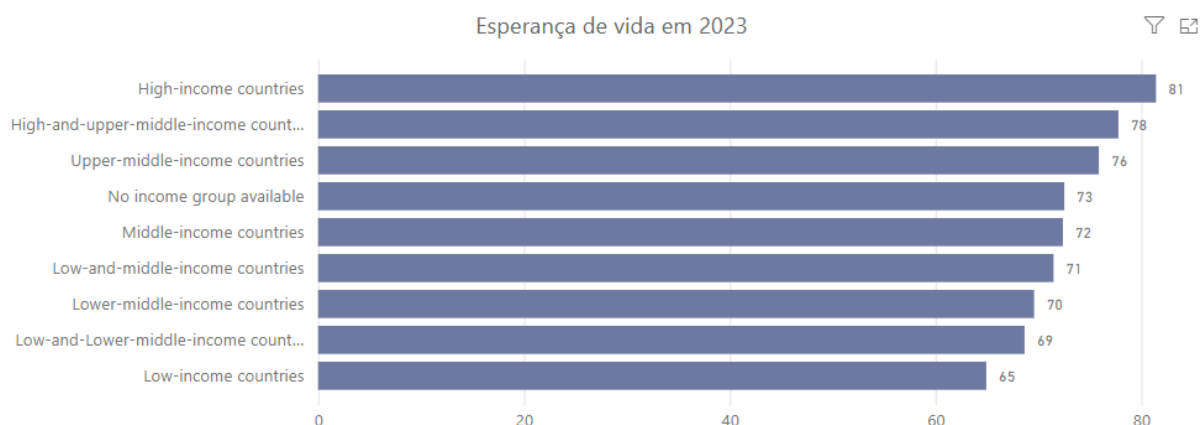


Figura 4.2 – EMV em 2023 por nível de rendimento dos países

Olhando agora para os países com a maior EMV em 2023, o TOP 10 é composto pelos países que constam no Quadro 4.1, e como se pode verificar, não se encontram no TOP países da América do Sul ou países africanos.

Quando analisados os dados de 1950, apenas um destes países constava no TOP 10, nomeadamente, São Marino.

Quadro 4.1 – Top 10 dos países com maior EMV em 2023

País	EMV em 2023	Continent
Monaco	86,37	Europe
San Marino	85,71	Europe
Hong Kong	85,51	Asia
Japan	84,71	Asia
South Korea	84,33	Asia
Saint Barthelemy	84,29	North America
French Polynesia	84,07	Oceania
Andorra	84,04	Europe
Switzerland	83,95	Europe
Australia	83,92	Oceania

Quando analisados os países com menor E.M.V em 2023 verificamos que apenas se encontram neste quadro países do continente africano, como é visível no Quadro 4.2.

Quadro 4.2 – Top 10 dos países com menor EMV em 2023

Burkina Faso	61,09	Africa
Benin	60,77	Africa
Guinea	60,74	Africa
Mali	60,44	Africa
Somalia	58,82	Africa
South Sudan	57,62	Africa
Central African Republic	57,41	Africa
Lesotho	57,37	Africa
Chad	55,07	Africa
Nigeria	54,46	Africa

Vamos agora fazer uma análise genérica da EMV por continentes. Começando pelo continente europeu verificamos que em 1950 a quase totalidade dos países no TOP 10 eram países nórdicos. Esta tendência inverteu-se e em 2023 a maioria dos países no TOP 10 são países do sul da Europa conforme se pode ver no Quadro 4.3.

De assinalar também que no TOP 3 se encontram países de pequena dimensão geográfica e com um número reduzido de habitantes (menos de 100.000).

Quadro 4.3 – Top 10 dos países com maior EMV em 2023 na Europa

País	EMV em 2023
Monaco	86,37
San Marino	85,71
Andorra	84,04
Switzerland	83,95
Italy	83,72
Spain	83,67
Liechtenstein	83,63
Gibraltar	83,55
France	83,33
Norway	83,31

Nos dados dos países asiáticos verificamos menos mudanças no TOP 10 entre o ano de 1950 e 2023.

A saída de alguns países do TOP 10 poderá estar relacionada com conflitos ou guerras que surgiram após a década de 50, nomeadamente, o Líbano e o Cazaquistão.

O Chipre está classificado como país asiático devido à sua localização geográfica a leste do Mediterrâneo, apesar de politicamente e culturalmente estar mais próximo da Europa.

Este é um dos países do TOP 10 conforme se pode ver no Quadro 4.4.

Quadro 4.4 – Top 10 dos países com maior EMV em 2023 na Ásia

País	EMV em 2023
Hong Kong	85,51
Japan	84,71
South Korea	84,33
Singapore	83,74
Macao	83,07
United Arab Emirates	82,91
Israel	82,41
Qatar	82,37
Cyprus	81,65
Bahrain	81,28

No Quadro 4.5 que dispõe o TOP 10 da EMV na Oceânia identifica-se um intervalo grande entre o país no primeiro lugar (Polinésia Francesa, 84,07) e o país em décimo lugar (Tonga, 72,89).

Esta diferença era muito menos significativa na Europa e na Ásia, mas a explicação poderá estar no menor número de países que integram a Oceânia.

Aliás, o vigésimo terceiro e último país na lista tem um valor significativamente baixo quando comparado com os dez primeiros (Nauru, 62,11).

Quadro 4.5 – Top 10 dos países com maior EMV em 2023 na Oceânia

País	EMV em 2023
French Polynesia	84,07
Australia	83,92
New Zealand	82,09
Northern Mariana Islands	78,81
New Caledonia	78,77
Wallis and Futuna	78,63
Guam	77,21
Tokelau	77,05
Cook Islands	75,28
Tonga	72,89

Uma análise visual dos mapas do continente africano mostra uma inversão do Sul para o Norte dos países onde se encontra a maior EMV entre os dados de 1950 e os de 2023.

No Quadro 4.6 identificamos Cabo Verde, Marrocos, Argélia, Tunísia e Egito no TOP 10. Em 1950 apenas Cabo Verde fazia parte do TOP 10.

Quadro 4.6 – Top 10 dos países com maior EMV em 2023 em África

País	EMV em 2023
Reunion	83,55
Saint Helena	76,79
Tunisia	76,51
Algeria	76,26
Cape Verde	76,06
Mayotte	76,05
Morocco	75,31
Mauritius	74,93
Seychelles	72,86
Egypt	71,63

No Continente norte-americano assinalamos novamente a grande quantidade de países “mais pequenos” no TOP 10, com exceção para o Canadá, conforme se pode ver no Quadro 4.7.

Curiosidade também para o facto dos Estados Unidos da América não se encontrarem no TOP 10, apesar de terem feito parte desse Top em 1950.

Quadro 4.7 – Top 10 dos países com maior EMV em 2023 na América do Norte

País	EMV em 2023
Saint Barthelemy	84,29
Canada	82,63
Martinique	82,57
Bermuda	82,31
Guadeloupe	82,05
Puerto Rico	81,69
Costa Rica	80,80
Cayman Islands	80,36
Saint Martin (French part)	80,22
Panama	79,59

Por último analisamos os dados do continente sul Americano e a primeira consideração a fazer vai para o fato de apenas o Chile ter uma EMV em 2023 acima dos 80 anos, conforme se pode ver no Quadro 4.8.

Todos os restantes países estão abaixo desse valor, algo que apenas se tinha visto para o continente africano.

Quadro 4.8 – Top 10 dos países com maior EMV em 2023 na América do Sul

País	EMV em 2023
Chile	81,17
Falkland Islands	79,21
Uruguay	78,14
Peru	77,74
Colombia	77,72
Argentina	77,39
Ecuador	77,39
French Guiana	76,98
Brazil	75,85
Paraguay	73,84

4.2 Visualização dos dados

Com recurso ao Python foram extraídas várias representações gráficas dos dados para uma interpretação mais rigorosa a fim de tirar outros “insights”, encontrar valores em falta, duplicados, outliers, etc.

Com vemos no histograma representado na Figura 4.3, a maior concentração da EMV nos diferentes países tem vindo a deslocar-se da esquerda (40 anos em 1950) para a direita (75 anos em 2023).

Também percebemos que ao longo dos tempos, a EMV nos países tem chegado a idades onde antes não chegava. Em 2023 há grandes concentrações de países com E.M.V acima dos 75 anos. Nos anos anteriores analisados, não havia praticamente nenhum país neste intervalo.

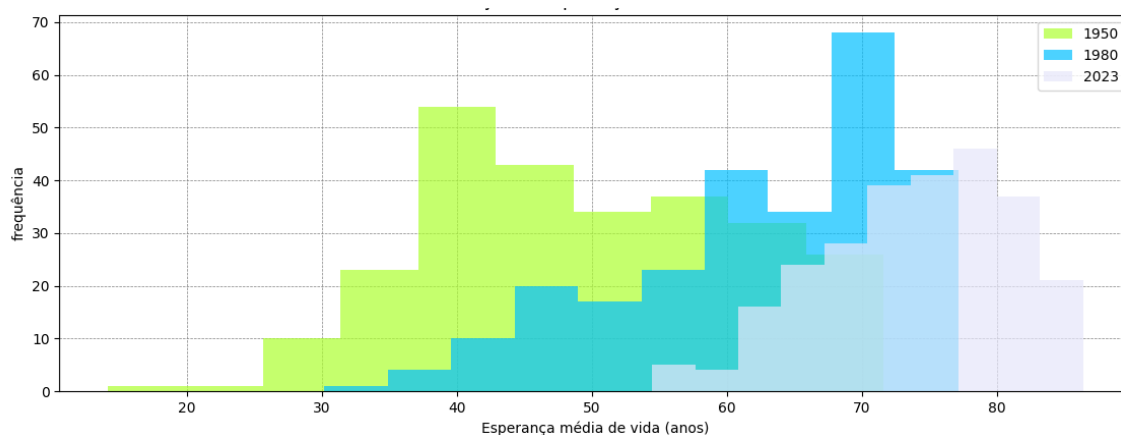


Figura 4.3 – Esperança média de vida global em 1950, 1980 e 2023

De seguida foram criados boxplots para cada um dos países para se procurarem identificar outras tendências que até aqui não tenham estado visíveis.

A Figura 4.4 apresenta os boxplots dos 15 países com a mediana mais elevada.

Da interpretação dos dados globais concluímos que os seguintes países têm outliers significativos, como por exemplo:

- Ilhas Faroé
- Japão
- Grécia
- Costa Rica
- Alemanha
- Áustria
- Cuba
- Sérvia
- Bulgária
- Ucrânia
- Coreia do Norte

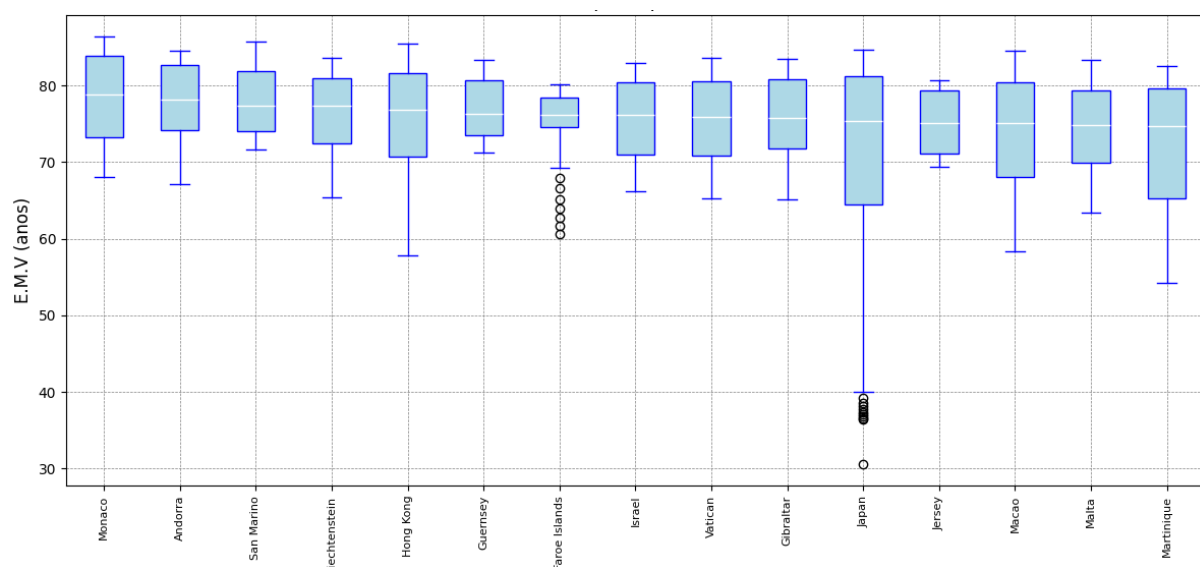


Figura 4.4 – Boxplot dos países com a mediana mais elevada

Concluiu-se que os boxplots tinham sido aplicados a todo o dataset e que os valores anteriores a 1950 aportavam muitos outliers aos boxplots, daí que se refez o exercício, mas desta vez apenas a partir de 1950.

Alguns exemplos de países com outliers são:

- Ilhas Faroé
- Montenegro
- Bahrain
- Sérvia
- Lituânia
- Bulgária
- Bielorrússia
- Dominica
- Estónia
- Letónia

- Roménia
- Gronelândia
- Iraque
- São Tomé e Príncipe
- Congo
- Quênia

Deste grupo, os países do Leste Europeu passaram por mudanças drásticas após o colapso da União Soviética e da Jugoslávia, o que pode explicar estes valores.

Outra análise efetuada foi relativamente aos países com grandes amplitudes entre valor mínimo e máximo.

Observam-se países onde os valores entre 1950 e 2023 variam entre 45 e 62 anos de diferença conforme se pode ver na figura 4.5.

Também se observam por esta leitura países que têm nos valores mínimos uma esperança de vida entre os 10 e os 20 anos, como por exemplo no Camboja, Sudão do Sul e Timor-Leste.

Em sentido contrário foram analisados os países com menores amplitudes entre o valor mínimo e máximo.

Aqui encontramos países que em 73 anos têm uma variação entre os 6 e os 12 anos, ou seja, não acompanharam a tendência global do aumento da esperança média de vida.

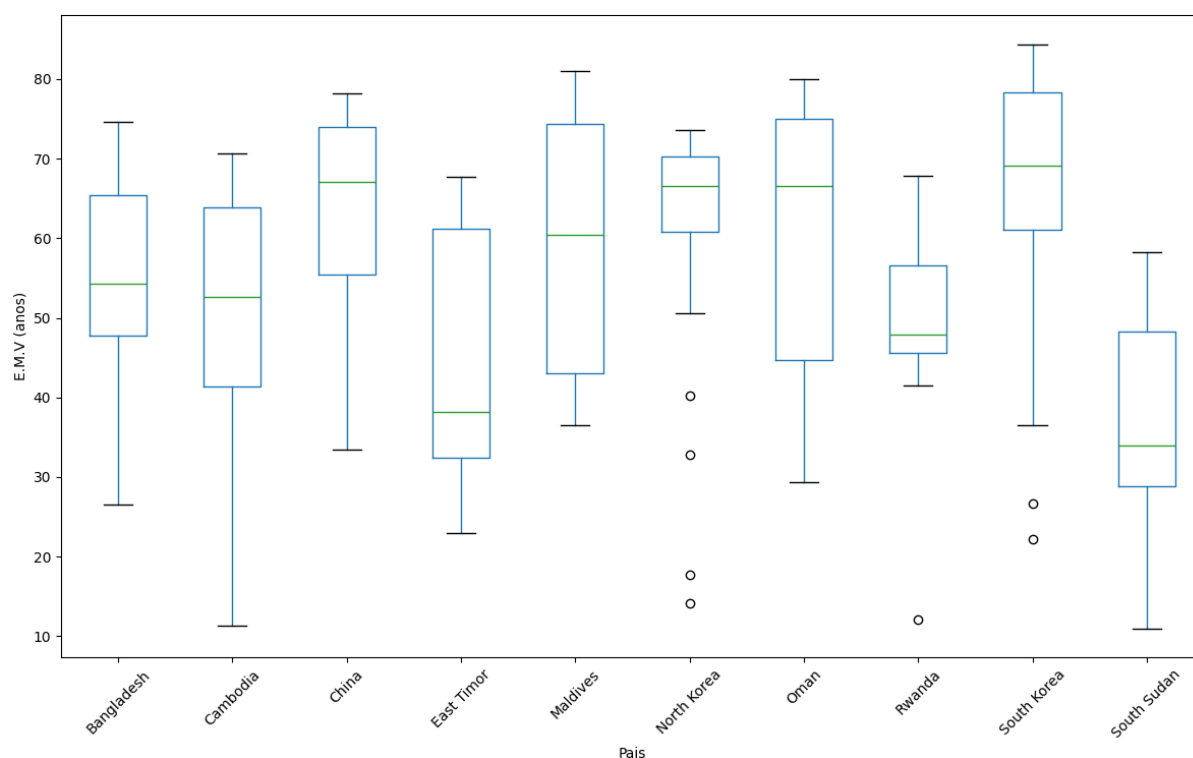


Figura 4.5 – Top 10 dos países com maior amplitude de valores entre 1950 e 2023

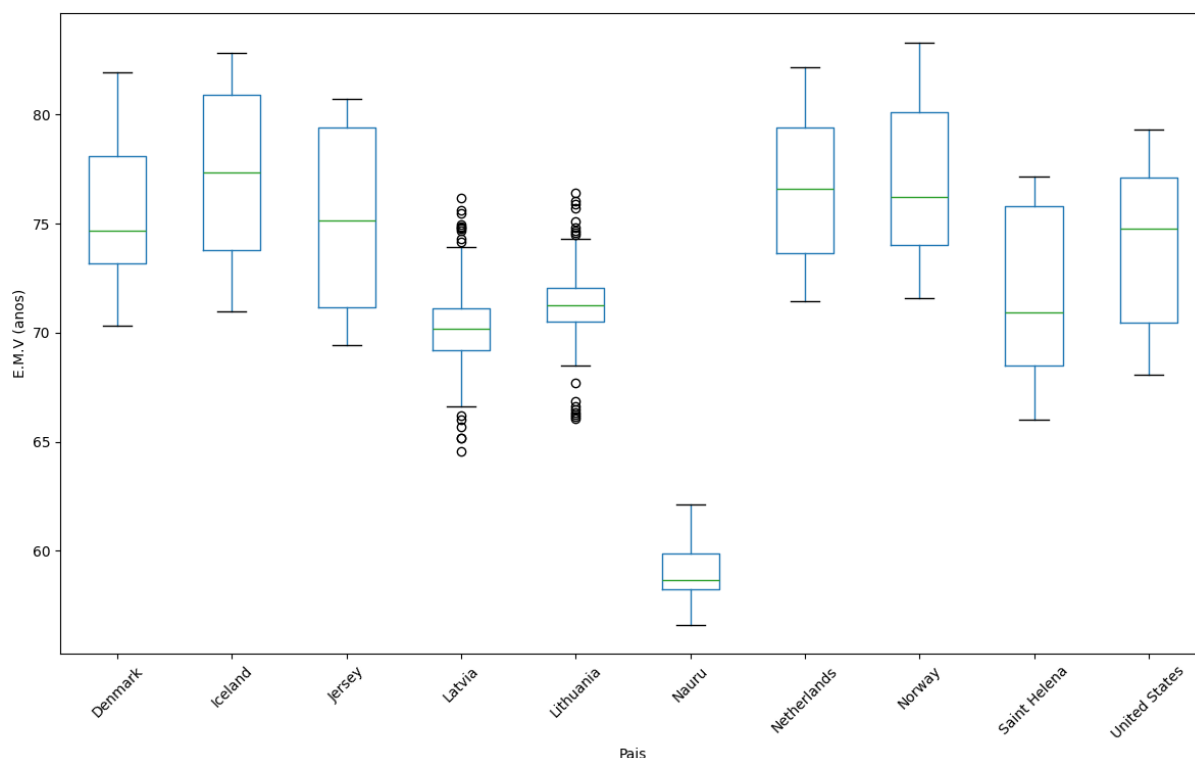


Figura 4.6– Top 10 dos países com menor amplitude de valores entre 1950 e 2023

Outra forma de olhar para os dados é através de mapas mundo onde podemos aplicar os dados de acordo com a informação geográfica.

Desta forma podemos identificar padrões que seriam mais difíceis de perceber por outra forma de representação visual.

Fizemos essa análise com os dados da EMV em 1950 e em 2023 conforme pode ser visualizado na Figura 4.7.

Rapidamente percebemos que entre os dois mapas existe uma mudança drástica entre o padrão de cores mais à esquerda que representa valores da EMV mais baixos para um padrão de cores mais à direita da barra e que representa valores mais elevados da EMV

Também conseguimos facilmente identificar quais os países ou regiões onde a EMV é mais elevada.

A principal dificuldade neste tipo de visualização tem a ver com o facto de apenas se conseguirem identificar os países com maior dimensão geográfica, ou seja, a não ser que se ampliem os mapas de forma significativa, não é possível identificar no mapa os países que estão no Top 3 de 2023 (São Marino, Mónaco e Hong-Kong).

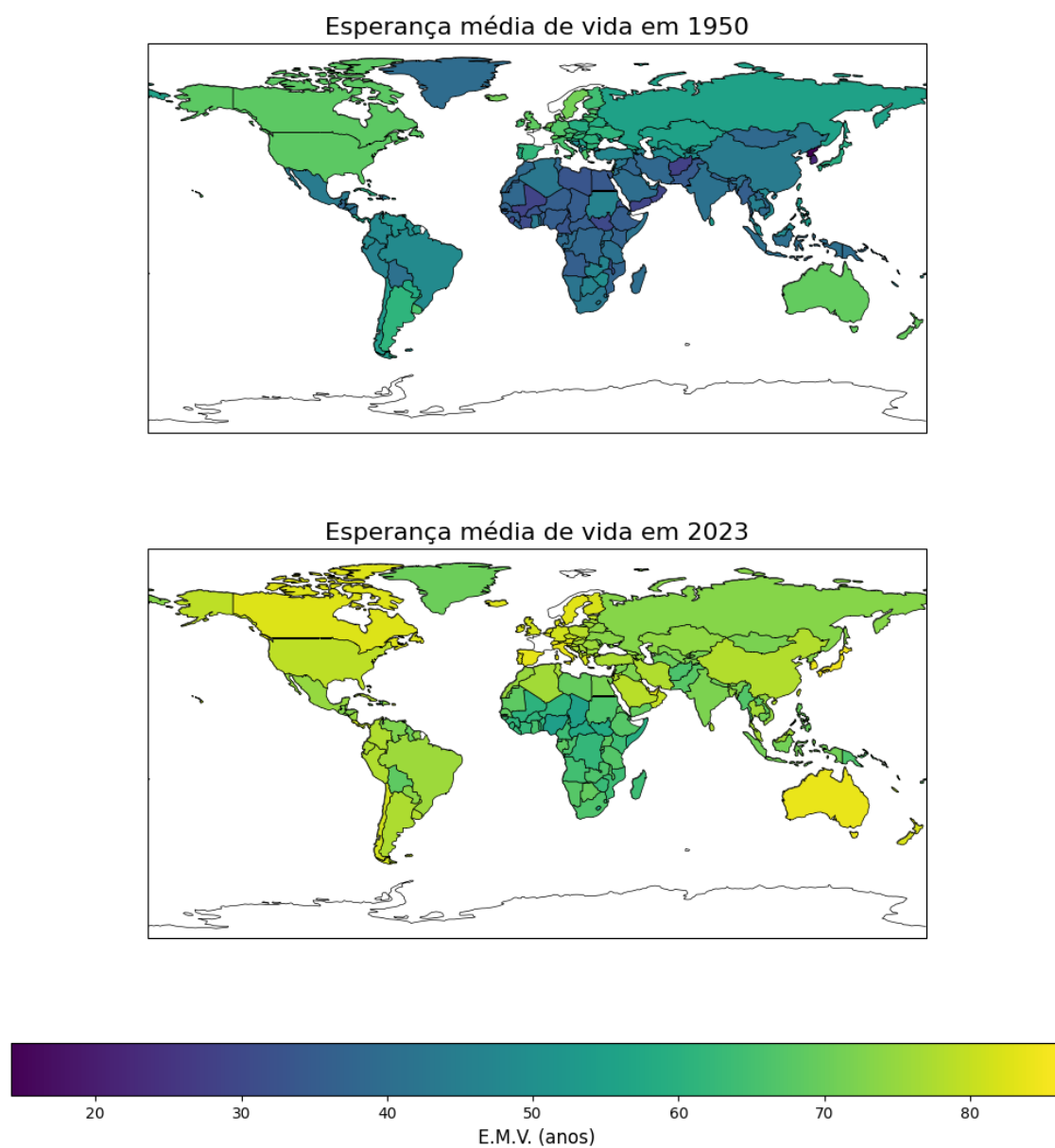


Figura 4.7 – Mapa mundo da EMV em 1950 e 2023

5. FATORES INFLUENCIADORES DA ESPERANÇA MÉDIA DE VIDA

Para esta análise foram importados os dados do website Our world in Data² divididos pelas seguintes categorias:

1. Violência e guerra
 - a. Mortes em conflitos armados
 - b. Direitos Humanos
 - c. Democracia eleitoral
2. Condições de vida, comunidade e bem-estar
 - a. Acesso a eletricidade
 - b. Acesso a água potável
 - c. Felicidade e satisfação com a vida
 - d. Taxa de desenvolvimento humano
3. Inovação e mudanças tecnológicas
 - a. Publicações científicas
 - b. Utilizadores de Internet
 - c. Transportes públicos
4. Educação e conhecimento
 - a. Anos de aprendizagem ajustados
 - b. Taxa de literacia
5. Pobreza e desenvolvimento económico
 - a. Pobreza extrema
 - b. PIB por capita
 - c. Gastos em saúde
 - d. Horas trabalhadas
6. Alimentação e Agricultura
 - a. Produção de cereais
 - b. Fornecimento de carne por pessoa
 - c. Consumo de carne por pessoa
 - d. Consumo de leite por pessoa
 - e. Consumo de ovos por pessoa
 - f. Consumo de peixe por pessoa
 - g. Utilização de fertilizantes por hectare
 - h. PIB agrícola
 - i. Subnutrição
 - j. Consumo calórico diário por pessoa
 - k. Obesidade
 - l. Fornecimento calórico diário
 - m. Consumo de frutas por pessoa
 - n. Consumo de vegetais por pessoa
 - o. Altura média

² <https://ourworldindata.org>

7. Energia e meio ambiente
 - a. Emissões de CO2 por pessoa
 - b. Exposição a poluição do ar
 - c. Consumo de Energia por pessoa
8. Saúde
 - a. Taxa de suicídio
 - b. Doenças cardiovasculares
 - c. Consumo de álcool
 - d. Adultos fumadores
 - e. Consumo diário de cigarros
9. População e mudanças demográficas
 - a. População
 - b. Idade média
 - c. Taxa de género
 - d. Taxa de fertilidade
 - e. Taxa de mortalidade infantil
 - f. Taxa de população urbana
 - g. Número de emigrantes

5.1 Correlação individual dos fatores

Com recurso ao R, todos os datasets foram carregados para que se possa começar por fazer uma regressão linear entre a variável independente (esperança média de vida) e as variáveis dependentes (fatores identificados acima).

O projeto R foi identificado com o nome OE3.

Numa primeira abordagem, a ideia era colocar todas as variáveis num dataset único para trabalhar os dados a partir daí.

Como a dimensão dos dados é grande e era necessário aplicar técnicas para preencher os dados em falta, o tempo de computação tornava-se muito elevado.

Daí que se tenha optado depois por dividir os datasets de acordo com as categorias de cada variável.

De seguida foi necessário limpar registos errados e registos com data anterior a 1950.

Por fim foram preenchidas as células com dados em falta com recurso ao método knn que preenche os dados com valores baseados em observações similares.

Para cada uma das variáveis foi construído um gráfico de regressão linear simples que podem ser consultados no Apêndice 1 e cujos valores de R (coeficiente de correlação linear de Pearson) se representam na tabela seguinte:

Quadro 5.1 – Regressão linear simples de indicadores vários com EMV

Indicador	Correlação com EMV
Média de idades	0,71
Taxa de desenvolvimento humano	0,62
Proteína animal	0,6
Percentagem de adultos obesos	0,6
Altura média dos homens	0,57
Índice de democracia eleitoral	0,56
Gastos na saúde	0,56
Índice de direitos humanos	0,54
Disponibilidade de água potável	0,53
Acesso a eletricidade	0,49
Índice de felicidade e satisfação	0,44
Taxa de literacia	0,43
Anos de aprendizagem ajustados	0,42
Uso de fertilizantes por hectare	0,42
Proteína gorda	0,4
Pib per capita	0,39
Consumo calórico diário	0,39
Publicações científicas	0,37
Consumo de aves por habitante	0,31
Consumo de energia por habitante	0,31
Consumo de carne de porco por habitante	0,3
Consumo de leite por habitante	0,21
Proteína carboidratos	0,2
Consumo de vegetais por habitante	0,2
Acesso a transportes públicos	0,19
Quilogramas de carne por habitante	0,19
Consumo de ovos por habitante	0,18
Emissões de CO ₂	0,16
Consumo de peixe e marisco por habitante	0,14
Consumo de álcool	0,14
Percentagem de fumadores	0,11
Consumo de fruta por habitante	0,07
Número de utilizadores de internet	0,04
Produção de cereais	0,04
Proteína vegetal	0,03
Consumo diário de cigarros	0,03
Número total de emigrantes	0,03
Taxa de população	0,01
Consumo de carne de cabra e ovelha por habitante	-0,01
Percentagem feminina	-0,01
Doenças cardiovasculares	-0,04

Consumo de outras carnes por habitante	-0,06
Mortes em conflitos	-0,07
Poluição atmosférica	-0,15
Consumo de carne de vaca por habitante	-0,17
Taxa de suicídio	-0,17
Número de horas trabalhadas	-0,39
Percentagem de sub-nutrição	-0,39
Pib agrícola	-0,58
Pobreza extrema	-0,68
Taxa de fertilidade	-0,84
Mortalidade infantil	-0,87

Os indicadores onde existe maior relação positiva com a EMV estão no topo da tabela:

- Média de idades
- Taxa de desenvolvimento humano
- Proteína animal
- Percentagem de adultos obesos
- Altura média dos homens
- Índice de democracia eleitoral
- Gastos na saúde
- Índice de direitos humanos
- Disponibilidade de água potável

Contudo também existe grande relação, mas desta vez negativa com os seguintes indicadores:

- Pib agrícola
- Pobreza extrema
- Taxa de fertilidade
- Mortalidade infantil

5.2 Modelo de regressão linear múltipla

De seguida criou-se um novo projeto no R (OE3B), desta vez apenas com os dados referentes à Alimentação e Agricultura, para desta vez criar um modelo de regressão linear múltipla e compreender padrões relacionados com a alimentação e agricultura que possam afetar a EMV através do modelo preditivo criado.

Os modelos de regressão linear múltipla são os mais populares e permitem estabelecer a relação entre uma variável numérica dependente, Y com um conjunto de k variáveis independentes, $X_1, \dots, X_k$ através da seguinte fórmula:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

onde $\beta_0, \dots, \beta_k$ são os coeficientes de regressão e ϵ é o ruído ou parte inexplicável.

Neste tipo de análise avalia-se a performance do modelo, mas também se escolhem as variáveis independentes a incluir no modelo por forma a obter o modelo mais fiável possível.

No nosso caso a variável numérica dependente é a EMV e os restantes indicadores são as potenciais variáveis independentes.

Antes de criar o modelo fizemos uma análise visual para identificar se há multicolinearidade entre os indicadores a introduzir no modelo.

Verificamos que o nosso modelo tem duas variáveis vindas de datasets diferentes, mas cujo valor é o mesmo (fish.and.seafood e fish.and.seafood.kg.per.year.per.capita), e como tal, uma delas não irá entrar no modelo inicial.

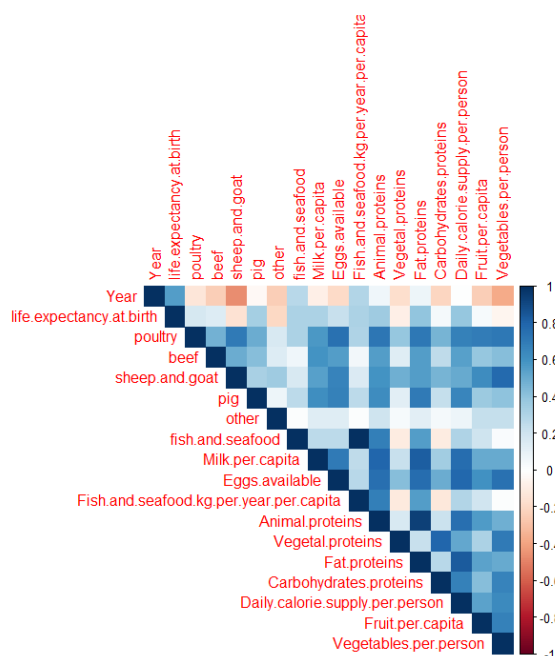


Figura 5.1 – mapa de multicolinearidade dos indicadores em estudo

Vamos procurar criar um modelo preditivo que permita prever novos valores o mais fiável possível e para tal, o dataset inicial foi dividido em dados de treino e dados de teste. Utilizamos 70% dos dados para treino e 30% para teste, para avaliar a capacidade do modelo se adaptar a novos dados.

Executando o modelo com os dados introduzidos obtemos os seguintes valores:

Quadro 5.2 – Regressão linear múltipla de indicadores vários com EMV

```

Residuals:
    Min       1Q   Median       3Q      Max
-48.779  -4.588   1.269   6.331  28.198

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    39.9818926   0.8186059   48.841 < 2e-16 ***
poultry         0.1038592   0.0106793    9.725 < 2e-16 ***
beef           -0.0569652   0.0131571   -4.330 1.51e-05 ***
sheep.and.goat -0.4133779   0.0128325  -32.213 < 2e-16 ***
pig            -0.0118702   0.0106578   -1.114 0.26541
other          -0.3572304   0.0252004  -14.176 < 2e-16 ***
fish.and.seafood 0.0177532   0.0087244    2.035 0.04188 *
Milk.per.capita 0.0100301   0.0024935    4.022 5.79e-05 ***
Eggs.available  0.1049007   0.0332822    3.152 0.00163 **
Animal.proteins 0.0225745   0.0040695    5.547 2.96e-08 ***
Vegetal.proteins -0.0410468   0.0044370   -9.251 < 2e-16 ***
Fat.proteins    0.0019992   0.0010018    1.996 0.04599 *
Carbohydrates.proteins 0.0003665   0.0009500    0.386 0.69966
Daily.calorie.supply.per.person 0.0105610   0.0007546   13.995 < 2e-16 ***
Fruit.per.capita -0.0155921   0.0020557   -7.585 3.57e-14 ***
Vegetables.per.person -0.0035795   0.0021658   -1.653 0.09841 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.183 on 12316 degrees of freedom
Multiple R-squared:  0.4201,    Adjusted R-squared:  0.4194
F-statistic: 594.8 on 15 and 12316 DF,  p-value: < 2.2e-16

```

O VIF é uma medida que indica o grau de multicolinearidade entre variáveis independentes.

Antes de proceder às conclusões vamos verificar os valores do VIF e identificar se há variáveis com VIF maior que 10, sugerindo multicolinearidade.

Quadro 5.3 – Valores de VIF dos indicadores no modelo

poultry	beef	sheep.and.goat	pig	other
4.547747	2.202942	4.877490	2.597870	1.384708
fish.and.seafood	Milk.per.capita	Eggs.available	Animal.proteins	Vegetal.proteins
6.003441	5.980408	5.489915	28.598458	4.619935
Fat.proteins	Carbohydrates.proteins	Daily.calorie.supply.per.person	Fruit.per.capita	Vegetables.per.person
18.613303	8.029570	17.890572	2.492414	7.284155

O valor dos indicadores Proteína animal (28.59), Proteína gorda (18.61) e Consumo diário de calorias (17.89) são maiores que 10, logo sugerem multicolinearidade e devem ser removidos do modelo.

Quadro 5.4 – Regressão linear múltipla excluindo valores indicados acima

```

Residuals:
    Min       1Q   Median       3Q      Max
-49.307  -4.873   1.280   6.462  42.035

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    47.1345541   0.7721896   61.040 < 2e-16 ***
poultry         0.1875959   0.0094959   19.756 < 2e-16 ***
beef            0.0490817   0.0122863    3.995 6.51e-05 ***
sheep.and.goat -0.4902239   0.0113131  -43.332 < 2e-16 ***
pig             0.0733680   0.0100246    7.319 2.66e-13 ***
other          -0.2583679   0.0241169  -10.713 < 2e-16 ***
fish.and.seafood 0.1106578   0.0043531   25.420 < 2e-16 ***
Milk.per.capita 0.0463573   0.0016779   27.628 < 2e-16 ***
Eggs.available  0.3002450   0.0321760    9.331 < 2e-16 ***
Vegetal.proteins -0.0145310   0.0043709   -3.324 0.000888 ***
Carbohydrates.proteins 0.0070506   0.0006257   11.269 < 2e-16 ***
Fruit.per.capita -0.0144308   0.0020853   -6.920 4.73e-12 ***
Vegetables.per.person -0.0104184   0.0021581   -4.828 1.40e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.394 on 12319 degrees of freedom
Multiple R-squared:  0.3929,    Adjusted R-squared:  0.3923
F-statistic: 664.4 on 12 and 12319 DF,  p-value: < 2.2e-16

```

Verificamos que o nosso valor de R^2 diminuiu para 0.3929 e o R^2 ajustado para 0.3923, mas o modelo ficou mais sólido, com um p-value extremamente baixo (2.2e-16), indicando uma significância estatística forte.

De seguida testamos o nosso modelo com os dados de teste com uma amostra de 20 registos.

O resultado foi o seguinte:

Quadro 5.5 – Previsão do modelo e verificação de exatidão

```

      Predição Atual Residual
3         60     29    -31.3
6         60     30    -30.0
8         60     31    -29.0
10        60     32    -28.0
14         57     34    -22.4
16         58     35    -22.5
22         55     38    -16.9
23         57     38    -18.1
24         57     39    -17.6
28         54     41    -13.2
29         55     40    -14.6
34         56     37    -19.0
38         56     39    -16.7
43         53     47     -6.1
47         54     53     -1.5
50         54     55     0.4
62         55     61     6.1
64         55     62     7.2
66         56     62     6.6
67         55     63     7.6
> options(scipen=999, digits=3)
> #calcular exatidão
> accuracy(emv.lm.pred,dados_teste$life expectancy.at.birth)
      ME RMSE MAE  MPE MAPE
Test set -0.0417 9.45 7.23 -3.25 13.3

```

Entre outros, podemos verificar que o Erro percentual absoluto médio é de 13.3%, o que para o caso é muito elevado.

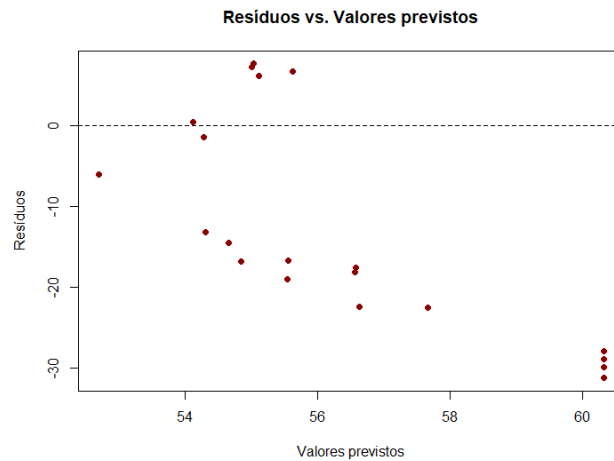


Figura 5.2 – visualização gráfica do teste efetuado

Vamos procurar otimizar o modelo fazendo uma busca exaustiva de variáveis (também chamada de best subset selection) para testar todas as combinações possíveis de variáveis e encontrar o modelo com melhor desempenho.

Através deste método concluímos que o modelo otimizado é o que já tínhamos inicialmente, ou seja, não sugere a retirada de nenhuma das variáveis do modelo, apesar de haver pouca melhoria a partir de 9 variáveis, e se quiséssemos simplificar o modelo deveríamos optar por usar apenas 9 variáveis.

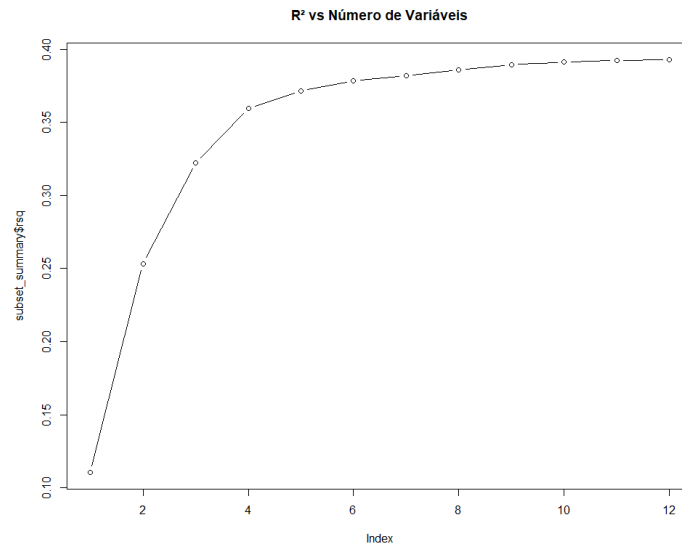


Figura 5.3 – visualização gráfica do método exaustivo

O mesmo acontece com o método alternativo que faz a busca forward, backward ou stepwise com base no critério AIC.

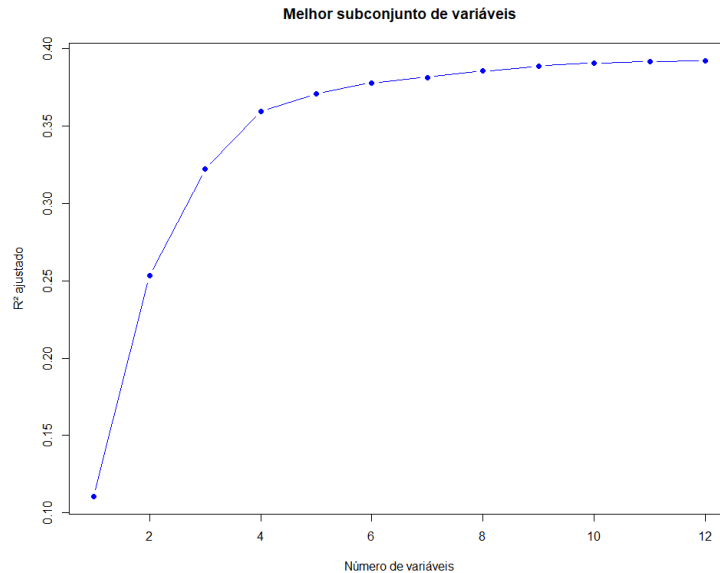


Figura 5.4 – visualização gráfica do método alternativo

Concluimos então que o modelo a que chegámos é o mais otimizado para um modelo de regressão linear múltipla. No entanto, a percentagem elevada de erro quando se pretende fazer uma previsão indica que com base nestas variáveis, não conseguimos de modo fidedigno fazer uma previsão da EMV.

5.3 Modelo de random forest

Este modelo é utilizado para casos em que a relação entre as variáveis não é linear e para o caso preparámos o projeto R com a designação OE3C.

Neste caso aplicamos um modelo de *Random Forest* para regressão usando 500 árvores e medindo a importância das variáveis.

Este modelo constrói várias árvores de decisão e combina os seus resultados para fazer previsões mais robustas.

Como é um modelo de regressão estamos a prever um valor contínuo, neste caso, a EMV.

Com 500 árvores de decisão temos estabilidade e bom desempenho do modelo.

A cada split (divisão num nó da árvore), o modelo considera aleatoriamente 5 variáveis das disponíveis. Isto introduz aleatoriedade e evita *overfitting*.

Obtivemos um valor de MSR (*Mean of squared residuals*), ou valor médio dos erros quadráticos entre os valores reais e os previstos) de 29.7.

Fazendo a sua raiz quadrada obtemos um valor de 5.45 que corresponderia ao RMSE e que indica um erro médio de previsão de cerca de 5.45 anos.

A percentagem de variância explicada é de 79.5%, o que indica que o modelo explica 79.5% da variância da esperança média de vida.

Assim sendo conclui-se que o modelo de *random forest* para estes dados em concreto é mais fiável que o modelo de regressão linear, já que o RMSE no caso desse modelo era de 9.45.

Fazendo ainda a interpretação da importância das variáveis verificamos que o gráfico abaixo indica a percentagem de aumento no erro quadrático médio no gráfico à esquerda e o aumento da pureza do nó à direita.

Quando maior a percentagem de IncMSE, mais importante é a variável para a previsão e quanto maior o IncNodePurity, mais importante é a variável.

Conclui-se que as variáveis mais importantes em ambas as métricas são: o consumo de ovos, o consumo de proteína gorda, o consumo diário de calorias, os outros e o consumo de proteína animal.

Por outro lado, as variáveis com pouco impacto são: o consumo de proteínas de carboidratos, o consumo de carne de vaca e o consumo de proteína vegetal.

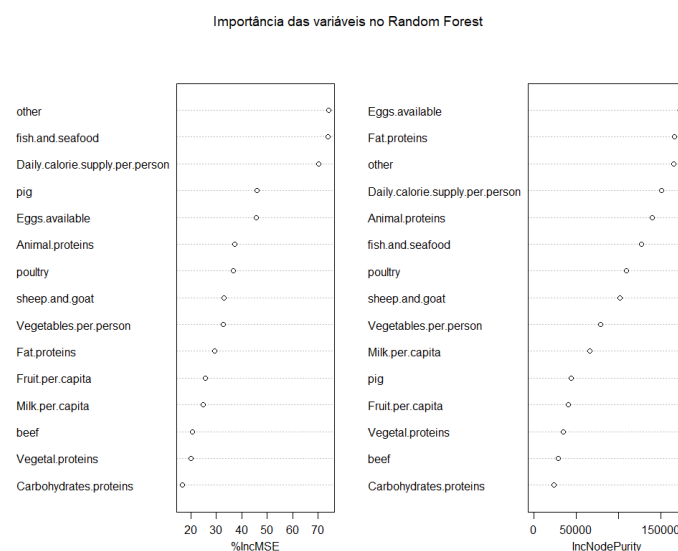


Figura 5.5 – importância das variáveis no modelo Random Forest

Vamos testar o modelo introduzindo primeiro dados arbitrários e verificar qual o output do modelo.

```
#testar com um dataset introduzido manualmente
novo_pais <- data.frame(
  Eggs.available = 10.2,
  Fat.proteins = 25.4,
  other = 15.6,
  Daily.calorie.supply.per.person = 2950,
  Animal.proteins = 18.7,
  fish.and.seafood = 8.3,
  poultry = 12.5,
  sheep.and.goat = 3.2,
  vegetables.per.person = 120,
  Milk.per.capita = 90,
  pig = 7.5,
  Fruit.per.capita = 100,
  vegetal.proteins = 20,
  beef = 5.6,
  carbohydrates.proteins = 60
)
```

Figura 5.6 – dados introduzidos para testar o modelo e verificar output

Com os dados introduzidos acima, o valor da EMV obtido foi de 62 anos.

Uma vez que as variáveis mais importantes são: o consumo de ovos, o consumo de proteína gorda, o consumo diário de calorias, os outros e o consumo de proteína animal, vamos aumentar o valor das mesmas e verificar o resultado obtido.

Aumentámos 100 unidades em cada indicador e o valor passou para 66.9 anos.

De seguida aumentámos 100 unidades em cada indicador nas variáveis com pouco impacto e o valor reduziu para 66.1 anos.

Verificamos assim a integridade dos indicadores com maior e menor impacto.

Por último testamos os dois modelos com dados para verificar qual se aproxima mais do valor real e o resultado é o apresentado na tabela abaixo.

Quadro 5.6 – Teste executado para os dois modelos

	Real	Regressao_Linear	Random_Forest	Erro_LM	Erro_RF
3	29.0	60.3	51.3	31.31	22.3208
6	30.4	60.3	51.3	29.96	20.9686
8	31.3	60.3	52.5	28.98	21.1119
10	32.3	60.3	52.5	27.99	20.1205
14	34.2	56.6	38.1	22.42	3.9001
16	35.1	57.7	36.6	22.54	1.5178
22	37.9	54.8	40.5	16.92	2.5901
23	38.4	56.6	41.2	18.13	2.7946
24	39.0	56.6	39.1	17.63	0.1614
28	41.1	54.3	42.2	13.22	1.0985
29	40.1	54.7	40.1	14.58	0.0304
34	36.5	55.5	42.5	19.02	5.9960
38	38.8	55.6	42.7	16.72	3.8342
43	46.6	52.7	49.2	6.13	2.6485
47	52.8	54.3	53.0	1.45	0.2159

Da análise da tabela concluímos que efetivamente o modelo de *random forest* é o mais adequado para este tipo de análise, já que o resultado do mesmo apresenta um erro inferior quando comparado com o modelo de regressão linear múltipla.

5.4 Modelo de regressão linear múltipla II

Por forma a consolidar o estudo, vamos repetir o mesmo exercício, mas desta vez com as variáveis que apresentaram maior correlação linear na abordagem da correlação individual dos fatores.

Assim, carregamos o nosso dataset com as seguintes variáveis:

- Média de idades
- Taxa de desenvolvimento humano
- Proteína animal
- Percentagem de adultos obesos
- Altura média dos homens
- Índice de democracia eleitoral
- Gastos na saúde
- Índice de direitos humanos
- Disponibilidade de água potável
- Pib agrícola
- Pobreza extrema
- Taxa de fertilidade
- Mortalidade infantil

Depois de fazer o respetivo tratamento dos dados, verificamos a multicolinearidade através do mapa gráfico.

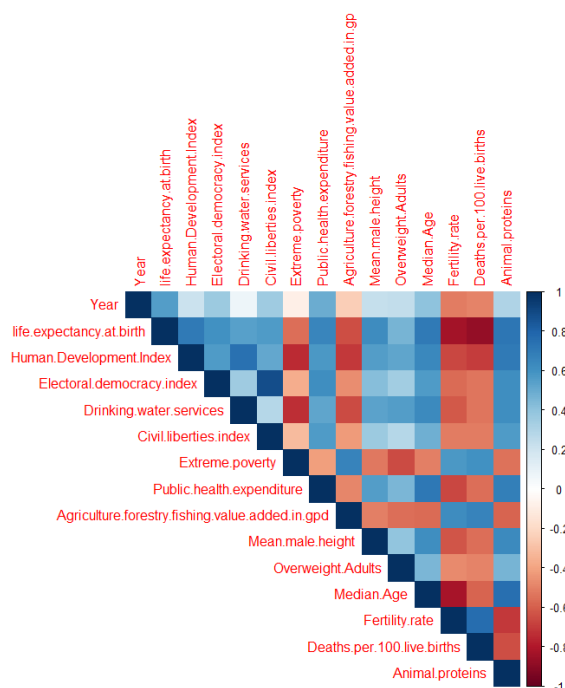


Figura 5.7 – mapa de multicolinearidade dos indicadores do segundo estudo

Os indicadores mais próximos são o índice de democracias eleitorais e o índice de liberdades civis, o que seria de esperar, já que estão intimamente relacionados.

Contudo vamos manter todas as variáveis no estudo e depois analisamos que valores nos dá a função vif.

Correndo o modelo de regressão linear múltipla obtemos o seguinte resultado e concluímos que o modelo bastante fiável, já que apresenta um R2 ajustado de 0,858 e um p-value extremamente baixo.

Quadro 5.7 – Regressão linear múltipla do segundo modelo com todas as variáveis

```
Residuals:
    Min       1Q   Median       3Q      Max
-39.89  -1.86   0.42   2.60  22.15

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  41.693770   2.252633   18.51 < 0.000000e+000 ***
Human.Development.index  5.548214   0.425357   13.04 < 0.000000e+000 ***
Electoral.democracy.index  1.089789   0.326976    3.33  0.00086 ***
Drinking.water.services -0.023938   0.002177  -10.99 < 0.000000e+000 ***
civil.liberties.index    0.540602   0.301843    1.79  0.07332
Extreme.poverty          0.031997   0.003227    9.92 < 0.000000e+000 ***
Public.health.expenditure  0.233736   0.034045    6.87  0.000000e+000 ***
Agriculture.forestry.fishing.value.added.in.gpd -0.004216   0.004573   -0.92  0.35656
Mean.male.height        0.162116   0.012798   12.67 < 0.000000e+000 ***
overweight.Adults       -0.004267   0.003124   -1.37  0.17202
Median.Age              0.047573   0.011048    4.31  0.0000167424855 ***
Fertility.rate          -1.695956   0.046628  -36.37 < 0.000000e+000 ***
Deaths.per.100.live.births -0.763216   0.009420  -81.02 < 0.000000e+000 ***
Animal.proteins         0.014015   0.000906   15.47 < 0.000000e+000 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.55 on 12318 degrees of freedom
Multiple R-squared:  0.858,    Adjusted R-squared:  0.858
F-statistic: 5.72e+03 on 13 and 12318 DF,  p-value: <0.000000e+000
```

Ainda assim, há variáveis com um p-value elevado, logo vamos ter de continuar a testar o modelo com a eventual remoção de variáveis.

Em relação à multicolinearidade, os valores obtidos para todas as variáveis são inferiores a 10, logo não existe multicolinearidade entre variáveis.

Quadro 5.8 – Valores de VIF dos indicadores no modelo II

Human.Development.index	Electoral.democracy.index	Drinking.water.services
4.22	5.23	3.41
Civil.liberties.index	Extreme.poverty	Public.health.expenditure
4.80	3.61	2.65
Agriculture.forestry.fishing.value.added.in.gpd	Mean.male.height	overweight.Adults
2.62	1.99	1.99
Median.Age	Fertility.rate	Deaths.per.100.live.births
4.76	5.31	3.49
Animal.proteins		
3.40		

Quando analisamos os resíduos em relação aos valores previstos podemos verificar que o Erro percentual absoluto médio é de 5,71%, logo um valor mais aceitável.

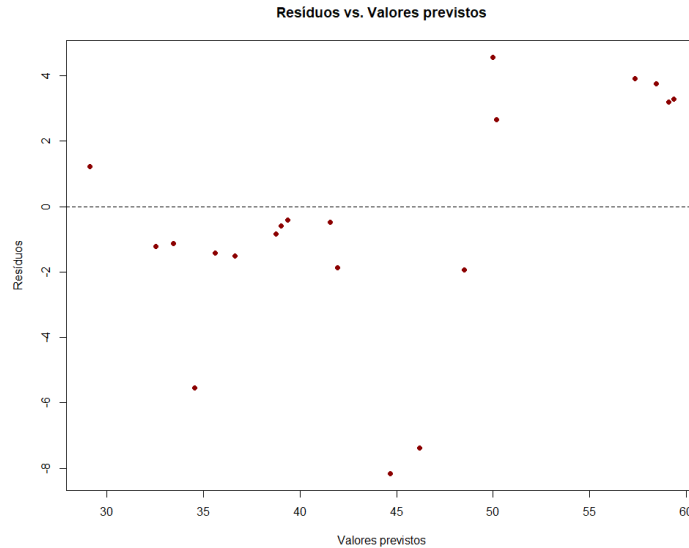


Figura 5.8 – visualização gráfica do teste efetuado com o modelo II

Vamos agora procurar otimizar o modelo removendo variáveis através do ajuste com busca exaustiva.

O resultado indica a retirada da variável referente ao PIB agrícola.

Através do método alternativo obtemos um modelo de 11 variáveis, em que são eliminadas as variáveis referentes ao PIB agrícola e de adultos com excesso de peso.

Neste caso obtemos um valor do R2 ajustado de 0,858 e um RSE de 4.55, ou seja, o mesmo que tínhamos no modelo inicial.

Quadro 5.8 – Regressão linear múltipla do modelo II ajustado com método alternativo

```

Residuals:
    Min       1Q   Median       3Q      Max
-39.88  -1.85    0.42    2.60   22.08

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   41.225025   2.232492   18.47 < 0.0000000000000002 ***
Human.Development.Index  5.670022   0.417505   13.58 < 0.0000000000000002 ***
Electoral.democracy.index  1.080658   0.326758    3.31  0.00094 ***
Drinking.water.services -0.023842   0.002137  -11.16 < 0.0000000000000002 ***
Civil.liberties.index    0.563253   0.301467    1.87  0.06173 .
Extreme.poverty          0.033370   0.002921   11.42 < 0.0000000000000002 ***
Public.health.expenditure 0.224786   0.033510    6.71  0.0000000000021 ***
Mean.male.height        0.163145   0.012780   12.77 < 0.0000000000000002 ***
Median.Age              0.047113   0.011036    4.27  0.00019767006 ***
Fertility.rate          -1.700611   0.046531  -36.55 < 0.0000000000000002 ***
Deaths.per.100.live.births -0.763667   0.009146  -83.49 < 0.0000000000000002 ***
Animal.proteins          0.014006   0.000906   15.46 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.55 on 12320 degrees of freedom
Multiple R-squared:  0.858,    Adjusted R-squared:  0.858
F-statistic: 6.75e+03 on 11 and 12320 DF,  p-value: <0.0000000000000002

```

5.5 Modelo de random forest II

Vamos agora aplicar o método de *random forest* ao mesmo conjunto inicial de dados e verificar se conseguimos obter melhores resultados que no método de regressão linear múltipla.

Obtivemos um valor de MSR (*Mean of squared residuals*), ou valor médio dos erros quadráticos entre os valores reais e os previstos de 3.3.

Fazendo a sua raiz quadrada obtemos um valor de 1,81 que corresponderia ao RMSE e que indica um erro médio de previsão de cerca de 1,81 anos.

A percentagem de variância explicada é de 97.7%, o que indica que o modelo explica 97.7% da variância da esperança média de vida.

Assim sendo conclui-se que o modelo de *random forest* para estes dados em concreto é mais fiável que o modelo de regressão linear e além disso, demonstra ser extremamente fiável.

Dada a importância das variáveis neste modelo, sugere-se uma nova iteração apenas com as variáveis mais importantes e verificar se a remoção das restantes afeta o desempenho do modelo.

Vamos manter apenas as variáveis Deaths.per.100.live.births, Fertility.rate, Median.Age, Human.Development.Index e Public.health.expenditure e verificar o resultado.

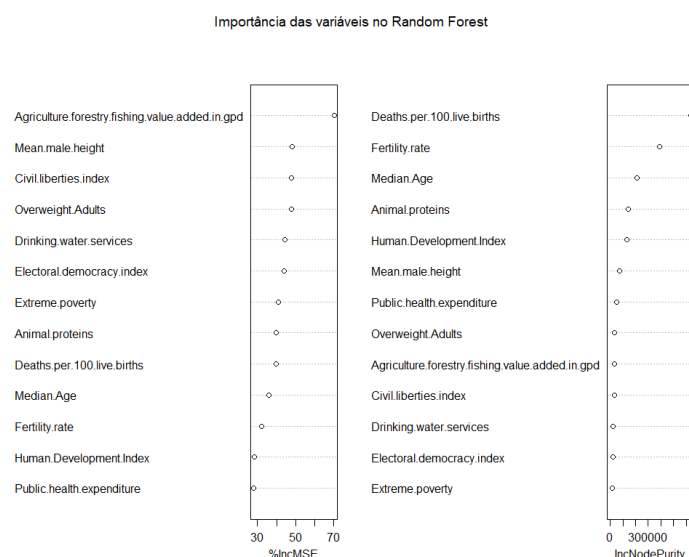


Figura 5.9 – importância das variáveis no Random Forest com o modelo II

Verificámos que apesar do modelo ficar mais simples, o seu desempenho é afetado já que o MSR aumentou para 7.2 e a percentagem de variância explicada caiu para 95.

Concluimos que o modelo ideal é o de *Random Forest* com todas as variáveis incluídas.

Fizemos um teste com dados introduzidos manualmente onde usámos os valores máximos ou mínimos de cada um dos indicadores e obtivemos uma EMV de 72.5 anos.

Também foi feito um comparativo entre os dois modelos e pudemos constatar que o Erro médio absoluto (MAE) é inferior para o caso do modelo *Random Forest*.

Quadro 5.9 – Comparativo dos modelos

```
> head(comparacao, 15)
      Real Regressao_Linear Random_Forest
3  29.0      34.7          34.1
6  30.4      29.3          32.9
8  31.3      32.6          33.4
10 32.3      33.5          35.0
14 34.2      35.7          35.4
16 35.1      36.7          36.2
22 37.9      38.8          38.5
23 38.4      39.1          39.3
24 39.0      39.5          40.3
28 41.1      41.5          40.6
29 40.1      41.9          40.2
34 36.5      44.6          39.0
38 38.8      46.3          42.5
43 46.6      48.6          50.2
47 52.8      50.1          51.8
>
> comparacao$Erro_LM <- abs(comparacao$Real - comparacao$Regressao_Linear)
> comparacao$Erro_RF <- abs(comparacao$Real - comparacao$Random_Forest)
>
> # Erro médio absoluto (MAE) de cada modelo
> mean(comparacao$Erro_LM)
[1] 3.15
> mean(comparacao$Erro_RF)
[1] 1.69
> |
```

5.6 Análise dos indicadores dos países Top 10 em 2023

Para esta análise foram carregados os dados de todos os países no R (projeto OE3D) e posteriormente filtrou-se para obter apenas os valores de 2023. Alguns indicadores não tinham valores em 2023 e não foram considerados.

De seguida identificaram-se os países com a maior esperança média de vida e foi criada uma coluna adicional para distinguir entre os países do TOP 10 e os restantes.

Posteriormente, para cada um dos indicadores foi criado um gráfico que tem na coordenada do X cada um dos indicadores e na coordenada do Y, a esperança média de vida.

Os países do TOP 10 estão identificados pelo nome e assinalados com uma cor diferente.

Todos os gráficos estão disponíveis no Apêndice 2.

Numa primeira fase foi feita uma análise a todos os gráficos para verificar em que indicadores os países do TOP 10 se encontravam maioritariamente num dos extremos do gráfico, indicando assim a importância desse indicador por uma questão de convergência.

Verificamos por exemplo que todos os países do TOP 10 têm 0 mortes por conflitos. No entanto o coeficiente de Pearson é baixo para este indicador ($R=-0,082$).

No que diz respeito ao acesso a água potável, com exceção do Japão, todos os restantes países do TOP 10 têm taxas elevadas de serviços de água potável e neste caso o R é de 0,65.

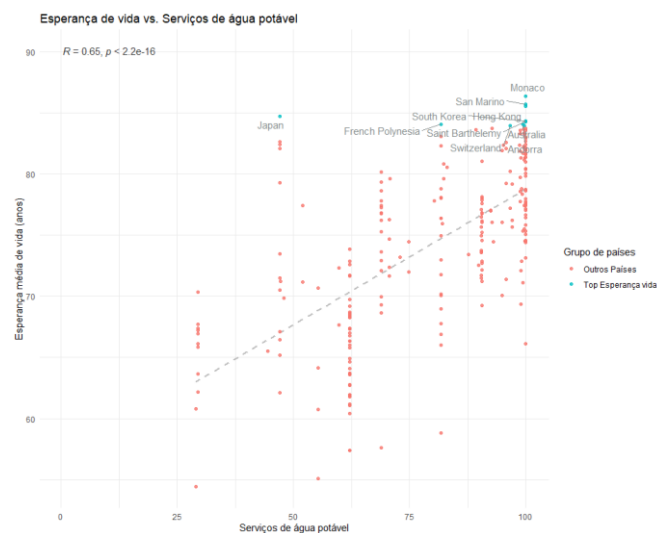


Figura 5.10 – esperança média de vida vs. Serviços de água potável

O índice de desenvolvimento humano também é elevado para todos os países no TOP 10, sendo o coeficiente de Pearson de 0,7 para este indicador.

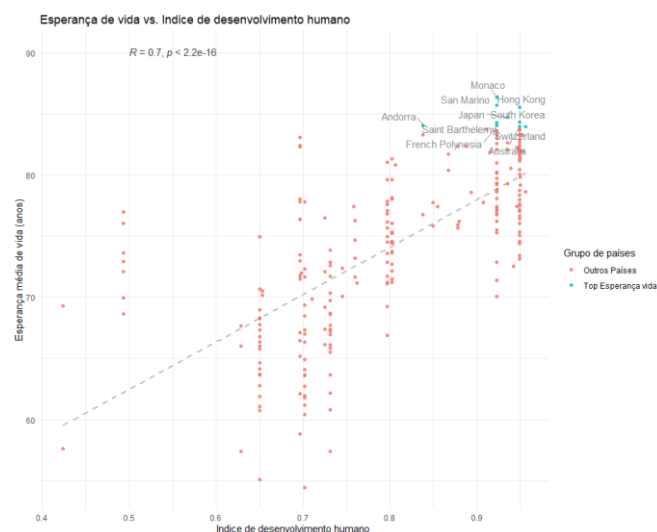


Figura 5.11 – esperança média de vida vs. Índice de desenvolvimento humano

O indicador de pobreza extrema coloca todos os países do TOP 10 na faixa dos países menos pobres, com exceção de Hong Kong. O R para este indicador é relativamente baixo (0,16).

Ao nível do PIB agrícola, todos os países no TOP tem valores baixos, com exceção da Polinésia Francesa.

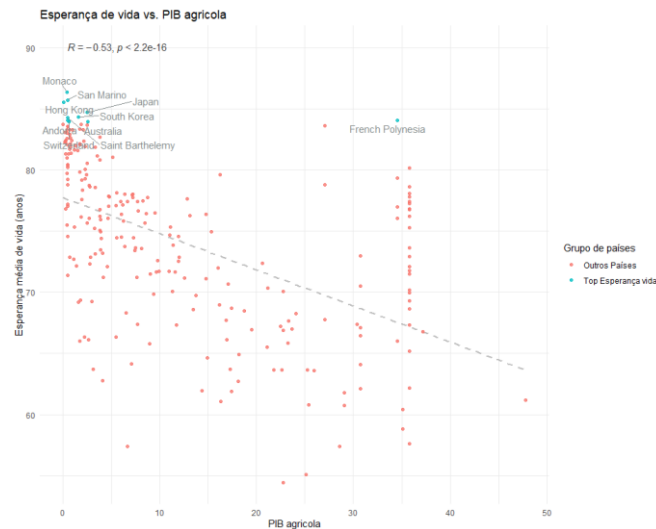


Figura 5.12 – esperança média de vida vs. PIB agrícola

Verificamos que no indicador de subnutrição, os países no TOP têm valores baixos de subnutrição, sendo que este indicador também tem um R considerável (-0,62).

Curioso é verificar que os países no TOP 10 têm níveis elevados de consumo de álcool, sendo o R de 0,65, ou seja, um valor significativo.

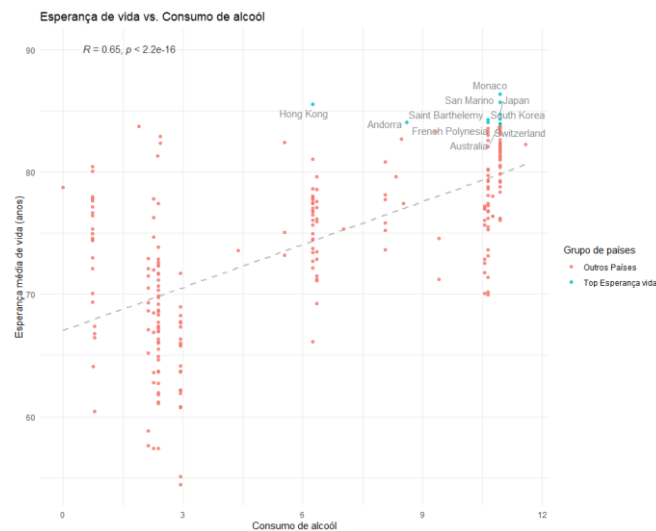


Figura 5.13 – esperança média de vida vs. Consumo de álcool

Por fim, dois dos indicadores com o coeficiente de Pearson mais elevado ($R=-0,8$ para índice de fertilidade e $R=-0,76$ para percentagem de mortes à nascença) indicam que os países no TOP 10 têm um baixo índice de fertilidade e baixos valores de mortes à nascença.

6. PREVISÃO DA EMV PARA OS PRÓXIMOS 50 ANOS

Um dos objetivos deste projeto é fazer uma projeção futura da EMV com base nos valores históricos de 1950 a 2023.

Nesse sentido foi criado o projeto em R, OE3E.

Para tal, usámos um modelo de regressão polinomial de 2º grau em que para cada país a curva é ajustada e com base nesse ajuste, é feita uma extrapolação para os valores futuros de 2023 a 2073.

Usando polinómios ortogonais, evitamos a colinearidade entre as variáveis ano e ano², o que melhora a estabilidade do modelo.

Ficam de seguida as projeções para alguns dos países mais significativos para este estudo:

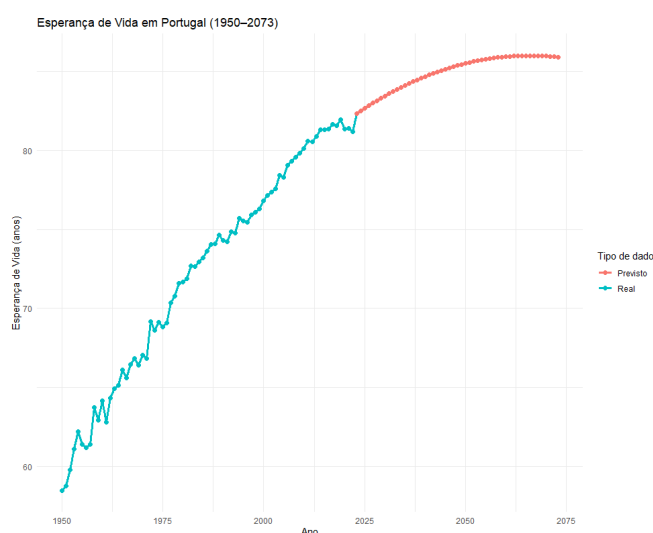


Figura 6.1 – EMV prevista para Portugal até 2073

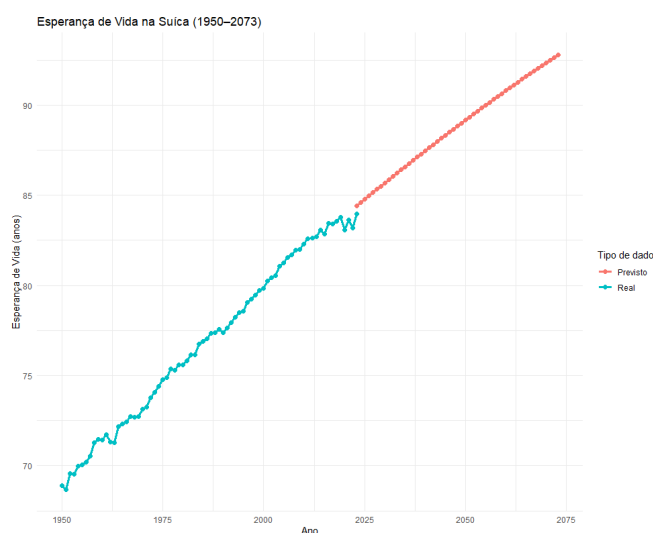


Figura 6.2 – EMV prevista para a Suíça até 2073

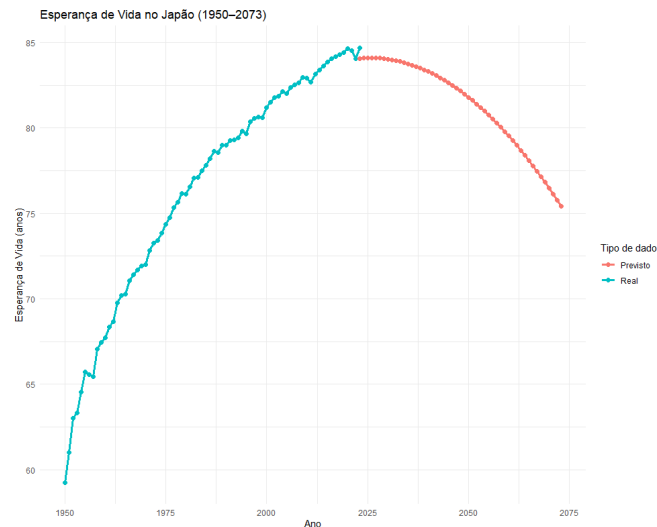


Figura 6.3 – EMV prevista para Japão até 2073

Considerou-se que estes países seriam significativos, uma vez que o primeiro é o país de origem do estudo e os outros dois são países com EMV alta, mas com curvas de evolução diferentes.

Comparativamente, foi usado outro método, neste caso um modelo logístico que impõe um limite natural superior que o modelo polinomial não faz.

O resultado para os mesmos países foi o seguinte:

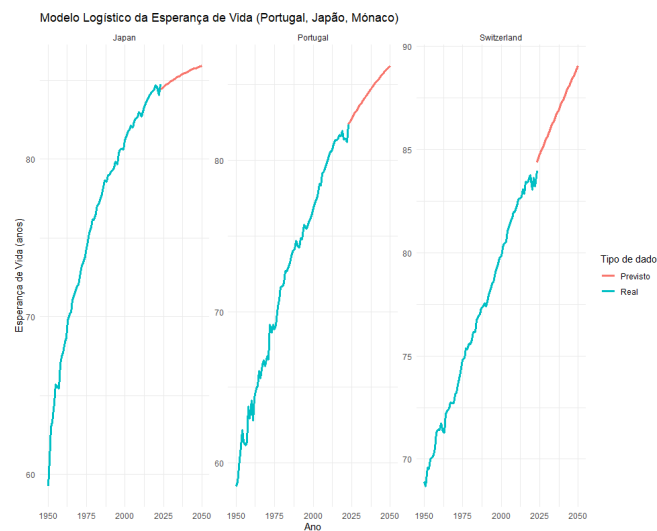


Figura 6.4 – EMV prevista para Japão, Portugal e Suíça até 2073, segundo método logístico

7. ANÁLISE REGIONAL DA ESPERANÇA MÉDIA DE VIDA EM PORTUGAL

7.1 Preparação

Para fazer uma análise espacial dos dados referentes a Portugal será utilizado o software ArcGis Pro, ferramenta que nos permite fazer uma análise temporal, de sobreposição ou estatística por forma a identificar padrões.

Apesar dos dados obtidos nas plataformas do INE ou do Pordata serem menores e menos detalhados do que os que conseguimos obter ao nível dos países, a representação visual será possível.

Criou-se um shapefile com os seguintes dados por região NUTSIII.

- Esperança de vida à nascença (Metodologia 2007 - Anos) por Local de residência (NUTS - 2024) - Anual (2021/2023)
- Taxa bruta de natalidade por Local de residência (NUTS - 2024) - Anual 2023-INE, Indicadores demográficos
- Taxa bruta de mortalidade por Local de residência (NUTS - 2024) - Anual 2023-INE, Indicadores demográficos
- Densidade populacional (N.º/km²) por Local de residência à data dos Censos [2021] (NUTS - 2024) e Sexo - Decenal - INE, Recenseamento da população e habitação - Censos 2021
- Índice de sustentabilidade potencial * (N.º) por Local de residência à data dos Censos [2021] (NUTS - 2024) e Sexo - Decenal - INE, Recenseamento da população e habitação - Censos 2021
- Distribuição do rendimento bruto declarado deduzido de IRS liquidado dos agregados fiscais por Localização geográfica (NUTS - 2024) e Quintis de rendimento - Anual 2022- INE, Estatísticas do Rendimento ao nível local com base na informação produzida pelo Ministério das Finanças - Autoridade Tributária e Aduaneira.

*Índice que mede a relação entre a população em idade ativa e a população idosa, representada habitualmente pelo quociente entre o número de pessoas com idades compreendidas entre os 15 e os 64 anos e o número de pessoas com 65 ou mais anos.

Estes dados estão agrupados pelas regiões das Entidades Intermunicipais (Comunidades Intermunicipais e Áreas Metropolitanas, no caso de Lisboa e Porto).

Para começar optou-se por usar um Shapefile com as regiões divididas pelos códigos NUTS (Nomenclatura das Unidades Territoriais para Fins Estatísticos) por forma a poder fazer posteriormente a junção das nossas tabelas com o Shapefile.

Um Shapefile é um formato de dados vectoriais para armazenar a localização, forma e atributos de feições geográficas.

7.2 Mapas temáticos

Assim foi possível criar mapas temáticos com os dados carregados no ArcGis Pro.

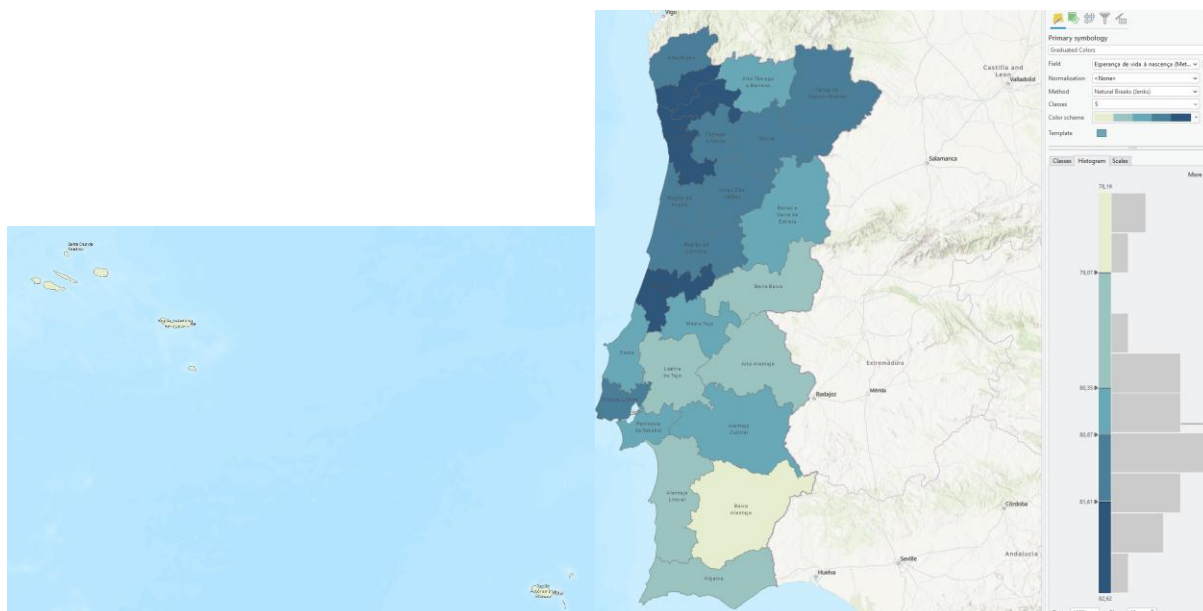


Figura 6.1 – esperança média de vida à nascença 2021-2023

Na figura acima percebemos que as regiões com a EMV mais alta para quem nasceu entre 2021 e 2023 são a região do Alentejo, Ave, Área Metropolitana do Porto e região de Leiria.

Já as regiões com os valores mais baixos são o Baixo Alentejo e as regiões autónomas dos Açores e da Madeira.

Uma forma de visualização que o ArcGis Pro permite é colocar dois mapas com indicadores diferentes, lado a lado, e desta forma facilitar a interpretação de quem está a analisar os dados.

Podemos ver essa análise nas figuras 6.2 a 6.6, onde a EMV à nascença entre os anos 2021 e 2023 por região no lado esquerdo da figura é comparada com os restantes indicadores em mapas posicionados no lado direito da figura.

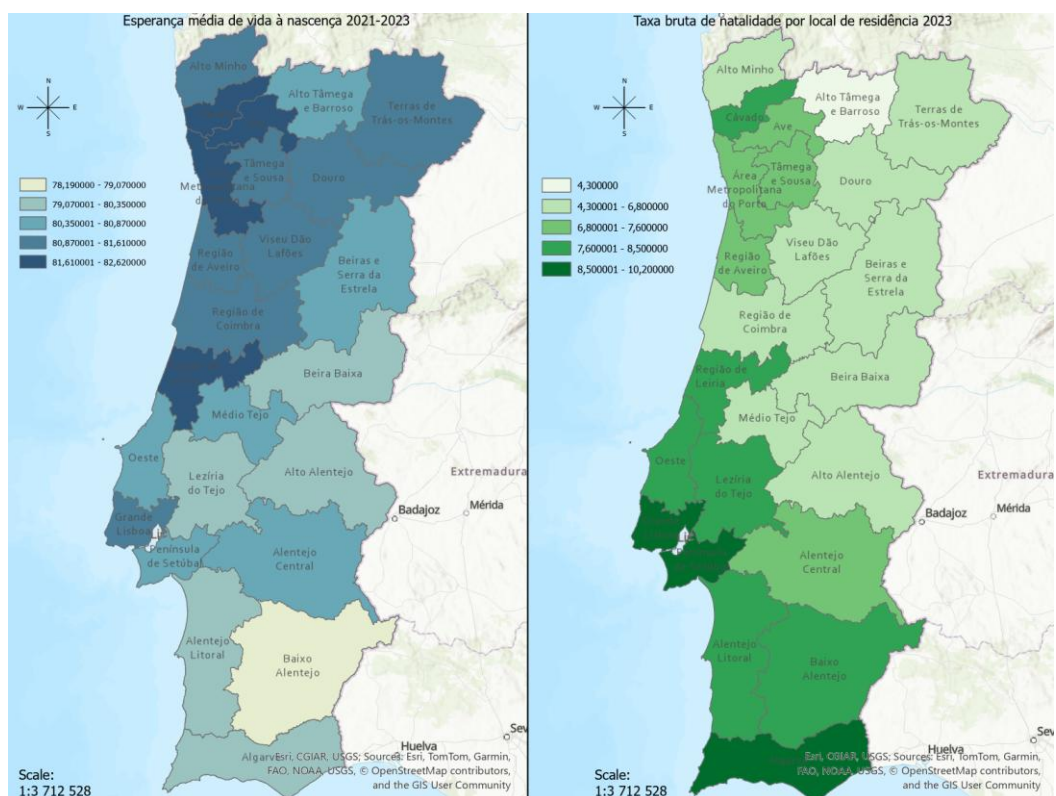


Figura 6.2 – esperança média de vida vs taxa de natalidade

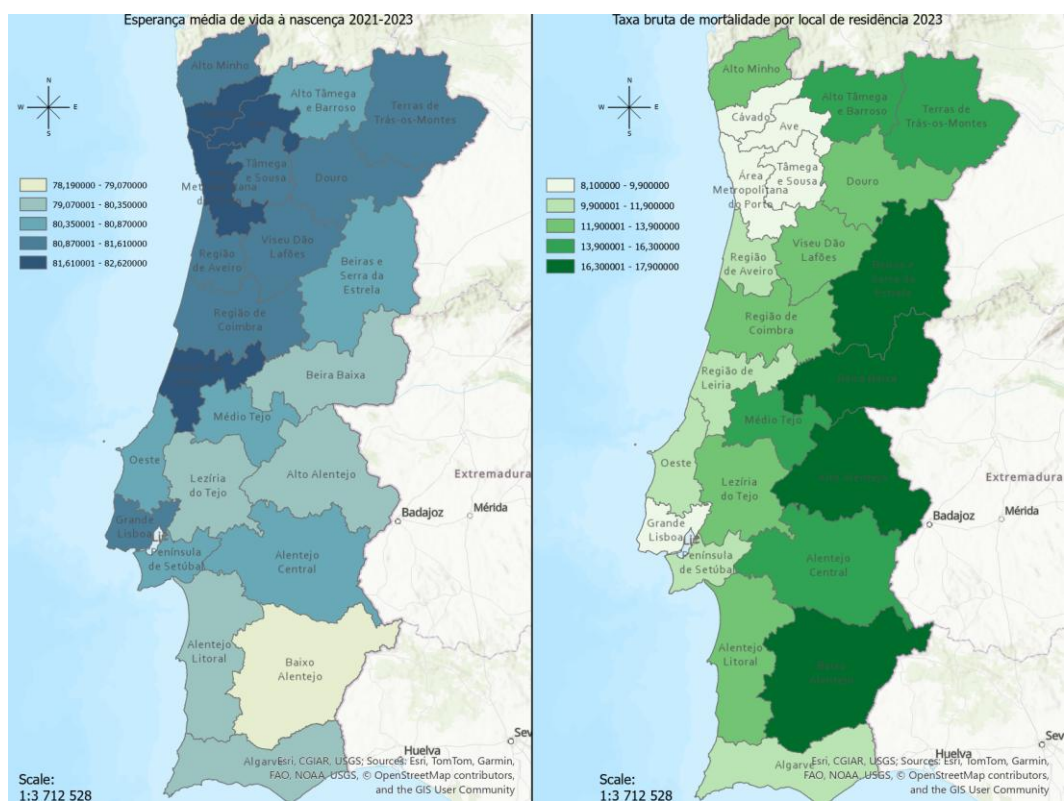


Figura 6.3 – esperança média de vida vs taxa de mortalidade

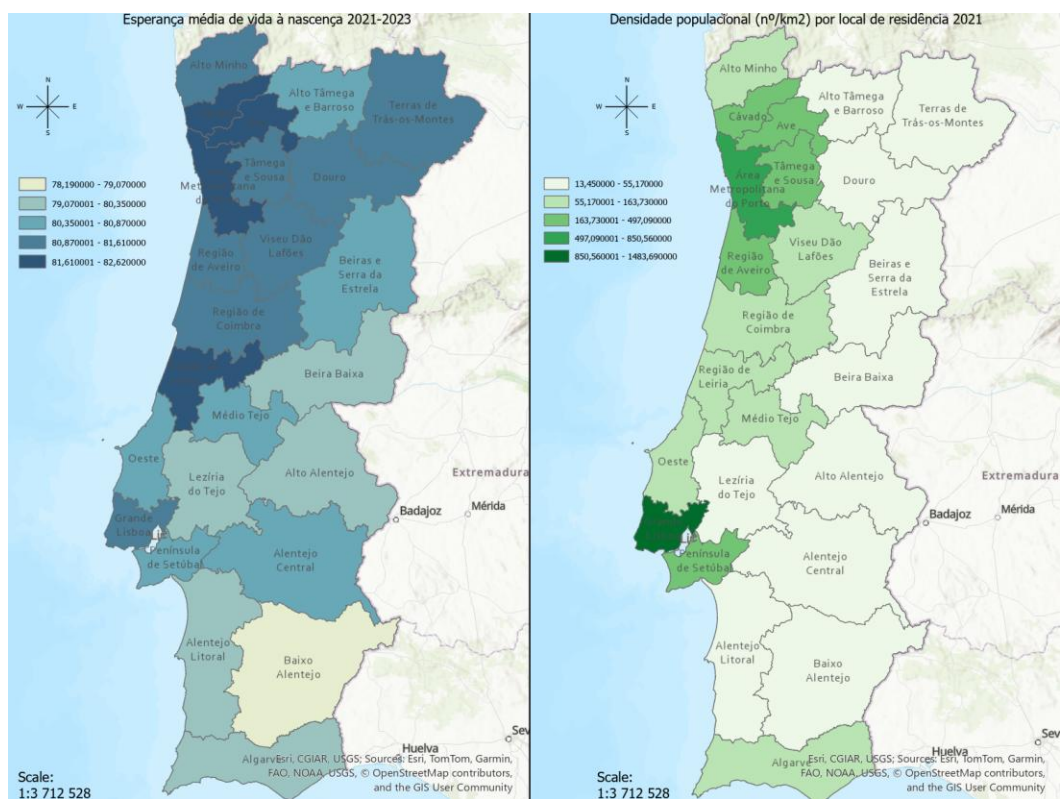


Figura 6.4 – esperança média de vida vs densidade populacional

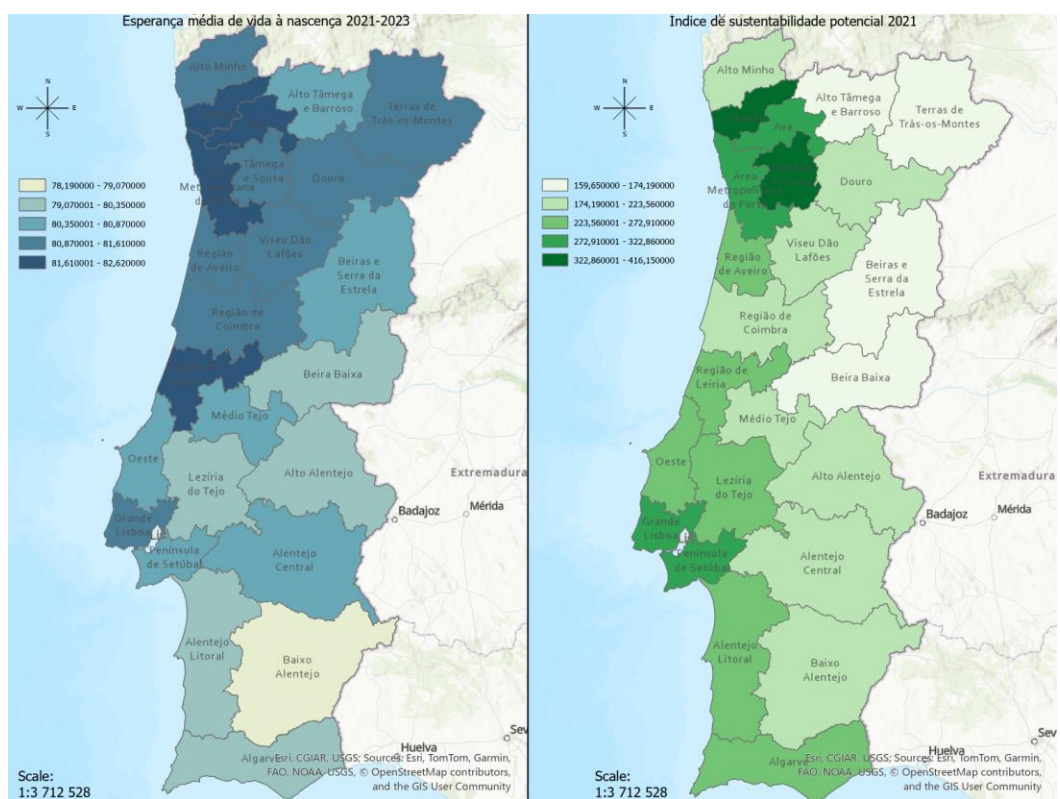


Figura 6.5 – esperança média de vida vs índice de sustentabilidade potencial

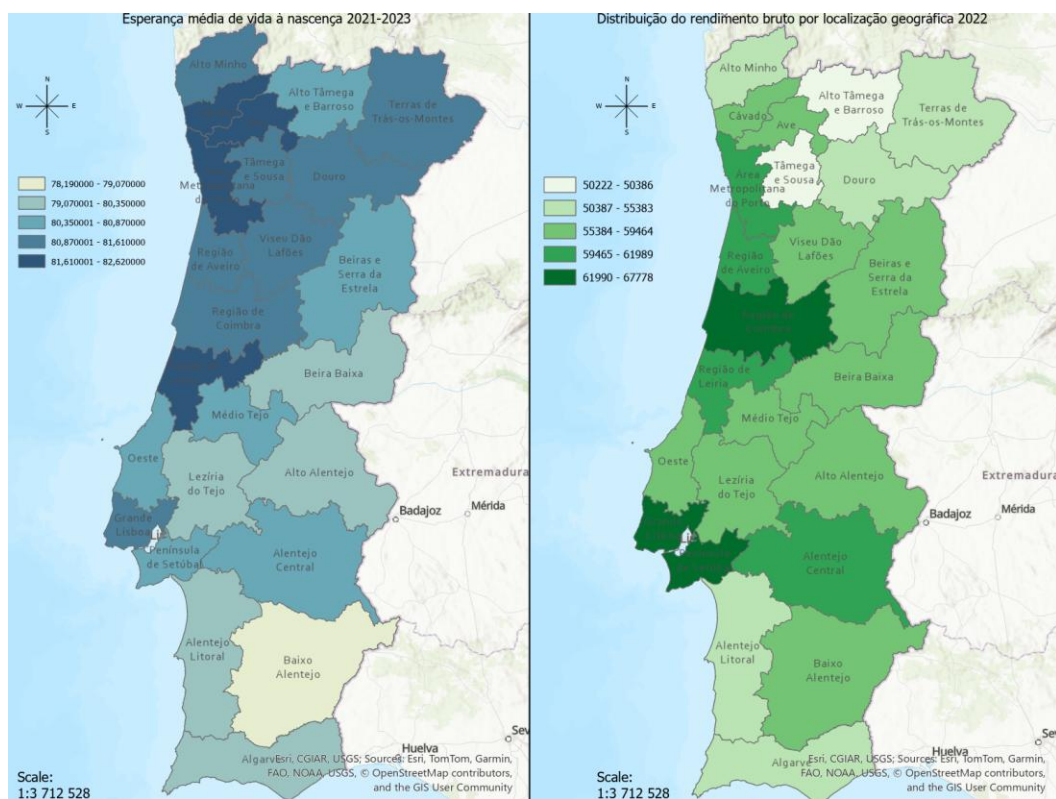


Figura 6.6 – esperança média de vida vs distribuição do rendimento bruto

Podemos concluir pela análise dos mapas comparativos que a taxa de natalidade aparenta ter alguma correlação com a EMV em Portugal, já que zonas como o Cávado, a Grande Lisboa, a zona de Leiria e toda a região litoral norte com exceção do Alto Minho apresentam taxa de natalidade e EMV elevada.

Por outro lado, o interior centro e sul, apresentam taxas elevadas de mortalidade e consequentemente, EMV mais baixas.

A densidade populacional concentra-se no Litoral a norte da Península de Setúbal e no Algarve.

De certa forma, também parece haver aqui alguma correlação entre os dois indicadores.

O índice de sustentabilidade potencial é aquele que parece ter mais influência sobre a esperança média de vida, uma vez que as manchas mais escuras de cada mapa se aproximam mais que nos anteriores.

Quanto à distribuição dos rendimentos, este parece ser o indicador com menor aproximação visual entre os dois mapas.

No entanto podemos usar a ferramenta Scatter plot matrix para aferir da relação entre as variáveis do nosso mapa.

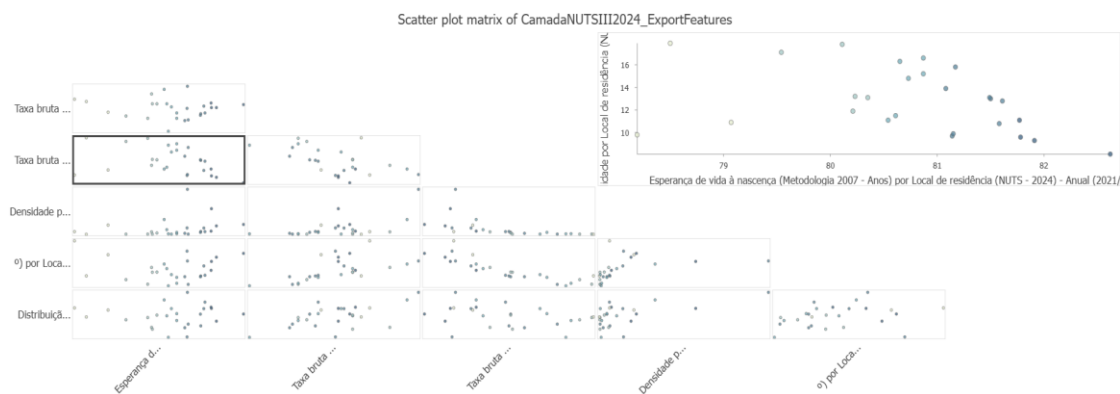


Figura 6.7 – scatter plot matrix para as variáveis em estudo

Pela análise do gráfico percebemos que existe pouca ou nenhuma relação entre a EMV por regiões e os restantes indicadores.

O que apresenta um R mais elevado (0,14) é a taxa de mortalidade.

As restantes variáveis apresentam valores iguais ou próximos de 0.

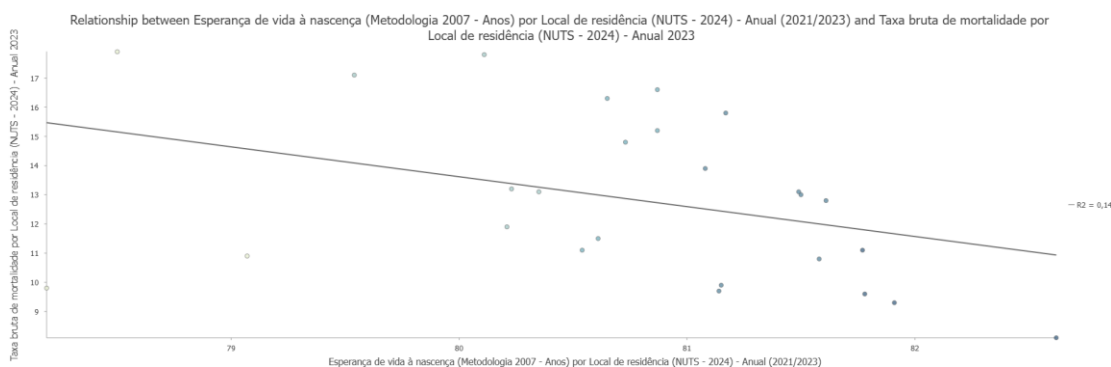


Figura 6.8 – scatter plot entre EMV e taxa de mortalidade

7.3 Dashboard interativo

Aproveitando o potencial do ArcGis, criou-se um “dashboard” interativo que permite neste caso identificar a relação entre a EMV e visualizar outros indicadores por região seleccionada.

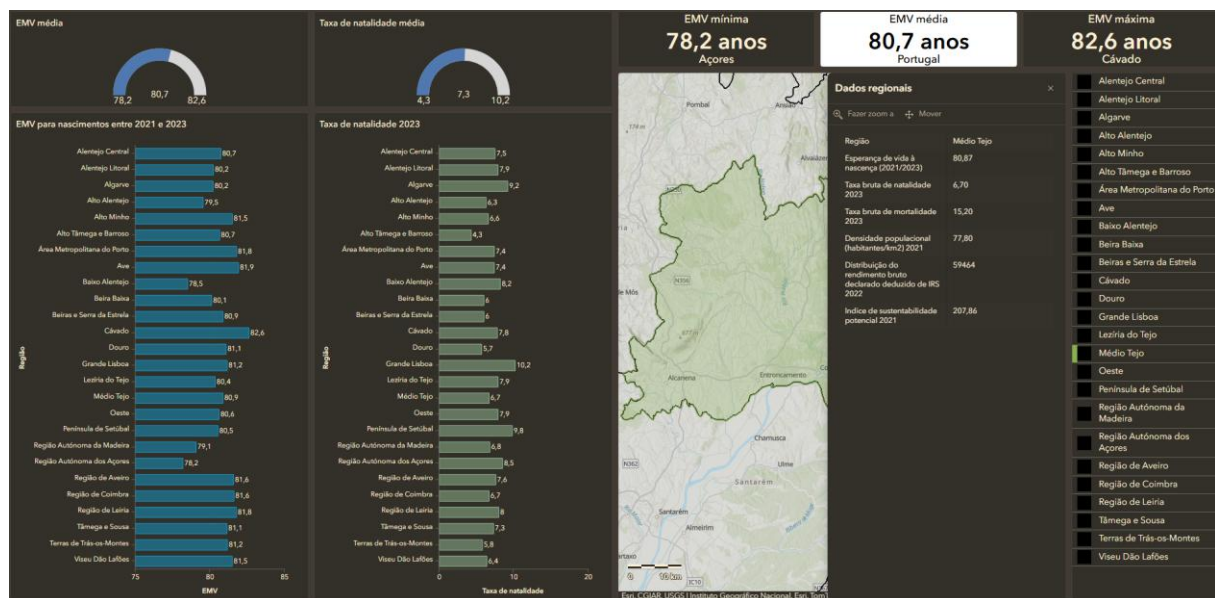


Figura 6.9 – Dashboard criado no ArcGis online

Neste dashboard podemos visualizar duas tabelas que apresentam os valores da EMV e taxa de natalidade para cada uma das regiões e dois medidores para os mesmos indicadores com o valor mínimo, médio e máximo.

Também temos o mapa de Portugal que permite, clicando numa determinada região, abrir uma janela com os indicadores dessa mesma região.

Também podemos escolher essa região na lista à direita, tendo acesso à mesma informação.

Link de acesso ao dashboard:

<https://www.arcgis.com/apps/dashboards/cea8352d3637476ca9a1cb37c3e6c633>

CONCLUSÃO E INVESTIGAÇÃO FUTURA

Após conclusão do projeto e com base nos dados analisados, foi possível verificar que de uma forma geral, a esperança média de vida tem vindo a aumentar de forma constante e regular, com exceção de períodos específicos como por exemplo os anos mais afetados pela pandemia da Covid 19.

Este aumento é mais significativo nos países com rendimento mais alto e com maior índice de desenvolvimento.

O continente africano é onde se registam os valores mais baixos de EMV, o que corrobora a informação anterior e reforça a ligação direta entre o desenvolvimento socioeconómico e a longevidade.

A análise da evolução temporal dos dados e a sua evolução e recorrendo a modelos de aprendizagem supervisionada, é de prever que a E.M.V continue a aumentar se não houver nenhum fator disruptivo de larga escala que possa contribuir para a inversão desta tendência.

Correlacionando os indicadores analisados com a EMV, há conclusões mais ou menos óbvias, como por exemplo o contributo positivo de indicadores como a Taxa de desenvolvimento humano, os gastos na saúde pública, a disponibilidade de água potável, o baixo índice de pobreza extrema ou mortalidade infantil.

Por outro lado, há indicadores que não são tão óbvios e que o estudo demonstrou haver maior EMV nos países onde esses indicadores são mais elevados.

Exemplo desses indicadores são a percentagem de adultos obesos e o consumo de álcool per capita.

Estes resultados, que podem refletir estilos de vida mais ocidentais ou fatores socioeconómicos indiretos, carecem de validação e investigação mais aprofundada.

No que respeita aos hábitos alimentares verificamos que o consumo de ovos, proteína gorda, consumo diário de calorias e consumo de proteína animal são indicadores importantes para a análise da EMV.

Pouco importantes são o consumo de carbo-hidratos, consumo de carne de vaca e consumo de proteína vegetal.

Outro aspeto interessante observado foi o perfil demográfico dos países com maior EMV em 2023. Verificou-se que estes países são, na sua maioria, estados de pequena dimensão ou territórios com baixa densidade populacional. Este padrão poderá indicar que fatores como urbanização controlada, qualidade ambiental ou políticas públicas mais ajustadas à escala populacional possam ter impacto positivo na longevidade — e, por isso, será uma linha de investigação relevante para trabalhos futuros.

Embora se tenha concluído que há indicadores que contribuem claramente para o aumento da EMV, há que reconhecer que a E.M.V resulta de um complexo sistema de interações entre estes e outros indicadores genéticos, comportamentais e sociais que tornam a captura de conclusões a partir de modelos lineares mais difíceis.

Analisando os dados nacionais, verificamos que em termos gerais, a EMV é mais alta a norte do que a Sul, com principal incidência nas zonas litorais.

A taxa de natalidade é mais alta no litoral Sul e a taxa de mortalidade no interior.

Ou seja, a E.M.V tem tendência a diminuir à medida que os habitantes mais velhos falecem e não se regista uma compensação por parte da natalidade que compense esse fator.

Para reforçar essa ideia podemos analisar a densidade populacional por região e verificar que todo o interior tem uma densidade populacional mais baixa que o litoral.

Conclui-se que se não forem tomadas medidas para compensar o envelhecimento e o êxodo populacional, haverá um agravar cada vez mais significativo da relação entre a taxa de mortalidade e a taxa de natalidade que levará a uma cada vez maior desertificação das regiões do interior do país.

Como nota final podemos confirmar que a quantidade e qualidade dos dados é essencial para a obtenção de resultados fidedignos que permitam a um Cientista de dados tratar os dados com diferentes abordagens para a obtenção de resultados que permitam fazer uma avaliação e análise eficaz e apresentar respostas ao problema inicialmente proposto.

Um dos indicadores que deverá ter impacto significativo na EMV é a taxa de tabagismo. Contudo, a escassez de dados relativamente a este indicador não permitiram tirar conclusões significativas sobre o mesmo.

A título pessoal posso concluir que este projeto permitiu cimentar os conhecimentos adquiridos no primeiro ano do mestrado e ir mais além, procurando outros métodos e abordagens que não foram abordadas nesse ano.

A escolha do tema relevou-se particularmente motivadora, uma vez que é um tema que me fascina e para o qual a quantidade de dados disponível permitiu ter uma abordagem não apenas focada numa determinada metodologia da Ciência de dados, mas a várias, que tornaram o processo mais estimulante.

Integrar este projeto com outras áreas sociais e económicas, poderá ajudar os decisores políticos a identificar as áreas com maior potencial de intervenção para contribuir para uma cada vez melhor qualidade de vida dos seus habitantes, reduzindo as assimetrias que foram verificadas nos diversos países ou regiões.

Será importante também ter em conta que o aumento da Esperança média de vida deverá ser acompanhada por um aumento da qualidade de vida.

O verdadeiro desafio para o futuro será não apenas prolongar a longevidade dos cidadãos, mas garantir que o aumento da Esperança média de vida traga também dignidade e saúde.

REFERÊNCIAS BIBLIOGRÁFICAS

- Martinez, I., Viles, E., & Olaizola, I. (2021). Data Science Methodologies: Current Challenges and Future Approaches.
- Abbasi-Kangevari, M., Malekpour, M. R., Masinaei, M., Moghaddam, S. S., Ghamari, S. H., Abbasi-Kangevari, Z., Rezaei, N., Rezaei, N., Mokdad, A. H., Naghavi, M., Larijani, B., Farzadfar, F., Murray, C. J. L., Saeedi Moghaddam, S., Aali, A., Abdollahzade, S., Abdulrahman, S. A., Abubaker Ali, H., Ahmad, A., ... Zareshahrabadi, Z. (2023). Effect of air pollution on disease burden, mortality, and life expectancy in North Africa and the Middle East: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Planetary Health*, 7(5), e358–e369. [https://doi.org/10.1016/S2542-5196\(23\)00053-0](https://doi.org/10.1016/S2542-5196(23)00053-0)
- Buxbaum, J. D., Chernew, M. E., Fendrick, A. M., & Cutler, D. M. (2020). Contributions of public health, pharmaceuticals, and other medical care to us life expectancy changes, 1990-2015. *Health Affairs*, 39(9), 1546–1556. <https://doi.org/10.1377/hlthaff.2020.00284>
- Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *JAMA - Journal of the American Medical Association*, 315(16), 1750–1766. <https://doi.org/10.1001/jama.2016.4226>
- Crimmins, E. M. (2015). Lifespan and healthspan: Past, present, and promise. *Gerontologist*, 55(6), 901–911. <https://doi.org/10.1093/geront/gnv130>
- Fadnes, L. T., Celis-Morales, C., Økland, J. M., Parra-Soto, S., Livingstone, K. M., Ho, F. K., Pell, J. P., Balakrishna, R., Javadi Arjmand, E., Johansson, K. A., Haaland, Ø. A., & Mathers, J. C. (2023). Life expectancy can increase by up to 10 years following sustained shifts towards healthier diets in the United Kingdom. *Nature Food*, 4(11), 961–965. <https://doi.org/10.1038/s43016-023-00868-w>
- Gulland, A. (2016). Global life expectancy increases by five years. *BMJ (Clinical Research Ed.)*, 353, i2883. <https://doi.org/10.1136/bmj.i2883>
- Hao, L., Xu, X., Dupre, M. E., Guo, A., Zhang, X., Qiu, L., Zhao, Y., & Gu, D. (2020). Adequate access to healthcare and added life expectancy among older adults in China. *BMC Geriatrics*, 20(1). <https://doi.org/10.1186/s12877-020-01524-9>
- Klenk, J., Rapp, K., Büchele, G., Keil, U., & Weiland, S. K. (2007). Increasing life expectancy in Germany: Quantitative contributions from changes in age- and disease-specific mortality. *European Journal of Public Health*, 17(6), 587–592. <https://doi.org/10.1093/eurpub/ckm024>
- Kopec, J. A., Sayre, E. C., Shams, B., Li, L. C., Xie, H., Feehan, L. M., & Esdaile, J. M. (2022). The Impact of 51 Risk Factors on Life Expectancy in Canada: Findings from a New Risk Prediction Model Based on Data from the Global Burden of Disease Study. *International Journal of Environmental Research and Public Health*, 19(15). <https://doi.org/10.3390/ijerph19158958>

- Le Bourg, É. (2012). Forecasting continuously increasing life expectancy: What implications? In *Ageing Research Reviews* (Vol. 11, Issue 2, pp. 325–328). <https://doi.org/10.1016/j.arr.2012.01.002>
- Lo, W. C., Hu, T. H., Shih, C. Y., Lin, H. H., & Hwang, J. S. (2024). Impact of Healthy Lifestyle Factors on Life Expectancy and Lifetime Health Care Expenditure: Nationwide Cohort Study. *JMIR Public Health and Surveillance*, 10. <https://doi.org/10.2196/57045>
- Mathers, C. D., Stevens, G. A., Boerma, T., White, R. A., & Tobias, M. I. (2015). Causes of international increases in older age life expectancy. In *The Lancet* (Vol. 385, Issue 9967, pp. 540–548). Lancet Publishing Group. [https://doi.org/10.1016/S0140-6736\(14\)60569-9](https://doi.org/10.1016/S0140-6736(14)60569-9)
- Monsef, A., & Mehrjardi, A. S. (2015). Determinants of Life Expectancy: A Panel Data Approach. *Asian Economic and Financial Review*, 5(11), 1251–1257. <https://doi.org/10.18488/journal.aefr/2015.5.11/102.11.1251.1257>
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296 (5570), 1029-1031.
- Vallin, J. & Mesle F., (2010) POPULATION SOCIETIES & No. 473 D E C E M B E R 2 0 1 0. <http://www.demogr.mpg.de/>
- Reimers, C. D., Knapp, G., & Reimers, A. K. (2012). Does physical activity increase life expectancy? A review of the literature. In *Journal of Aging Research* (Vol. 2012). <https://doi.org/10.1155/2012/243958>
- Sugiura, Y., Ju, Y.-S., Yasuoka, J., & Jimba, M. (2010). Rapid increase in Japanese life expectancy after World War II. In *BioScience Trends* (Vol. 4, Issue 1). www.biosciencetrends.com
- Yari, F., Agheli, L., Sadeghi, H., & Dizaji, S. F. (n.d.). The socioeconomic and political determinants of life expectancy in selected countries of the world. *Social Determinants of Health*, 8(1), 2022. <https://doi.org/10.22037/sdh.v8i1.37839>

APÊNDICES

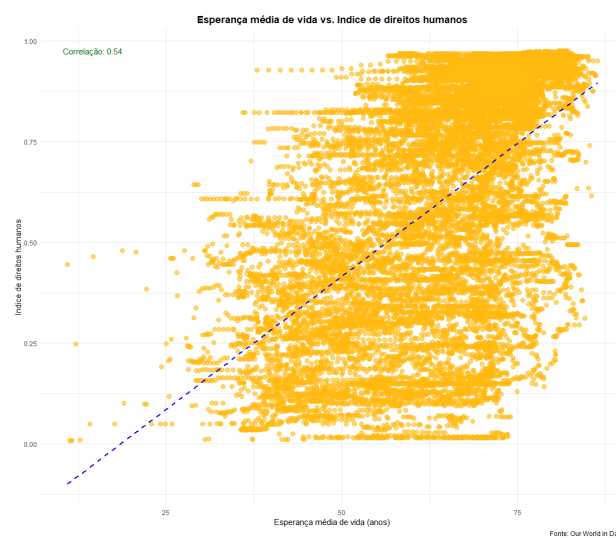
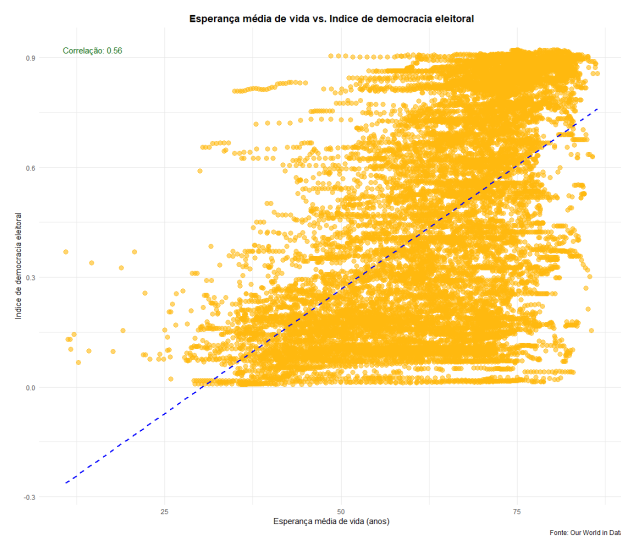
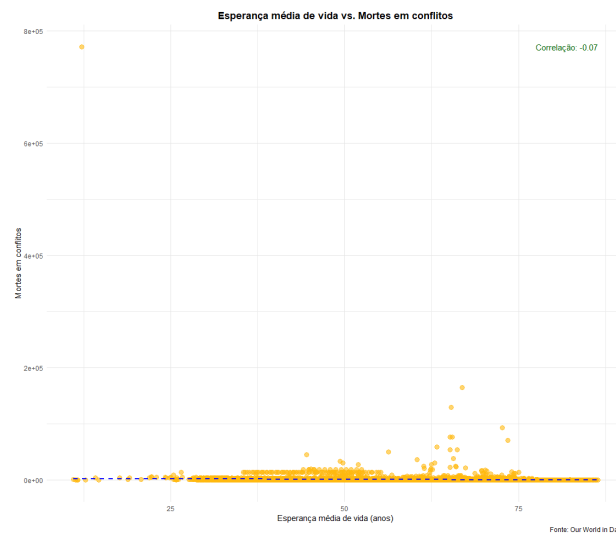
Apêndice 1 - Correlação linear entre esperança média de vida e outros fatores

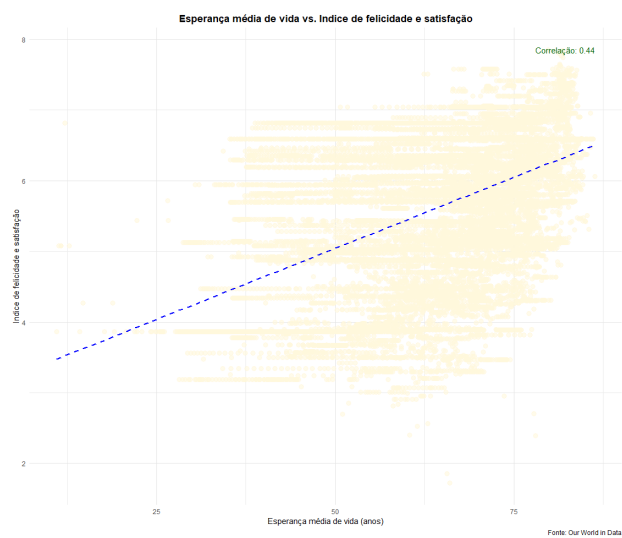
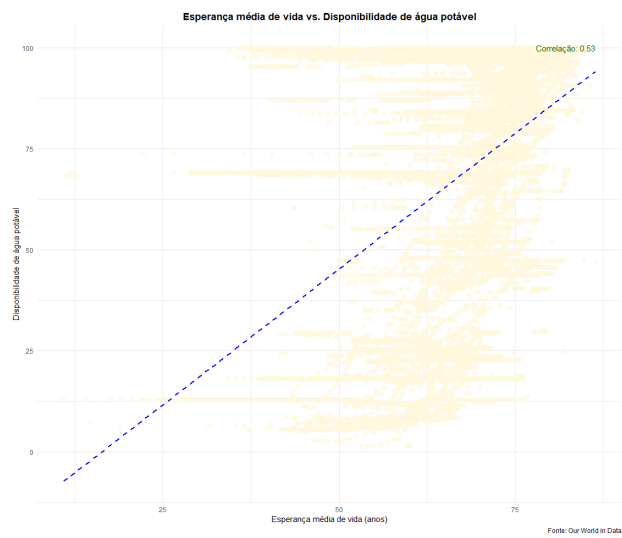
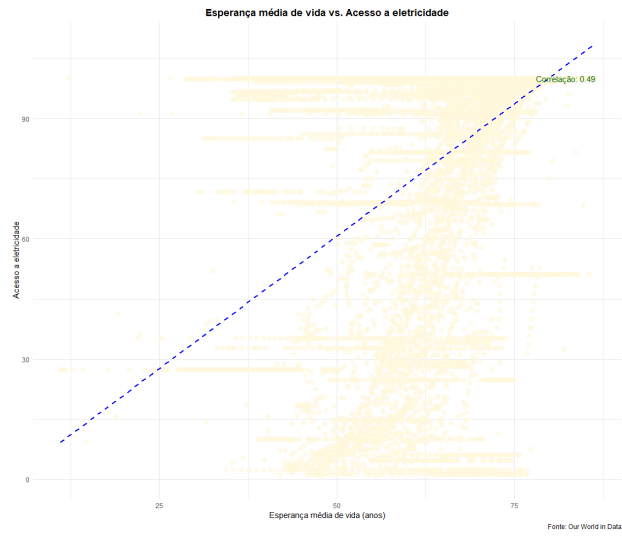
Apêndice 2 - Indicadores nos países Top Esperança média de vida vs. Outros países em 2023

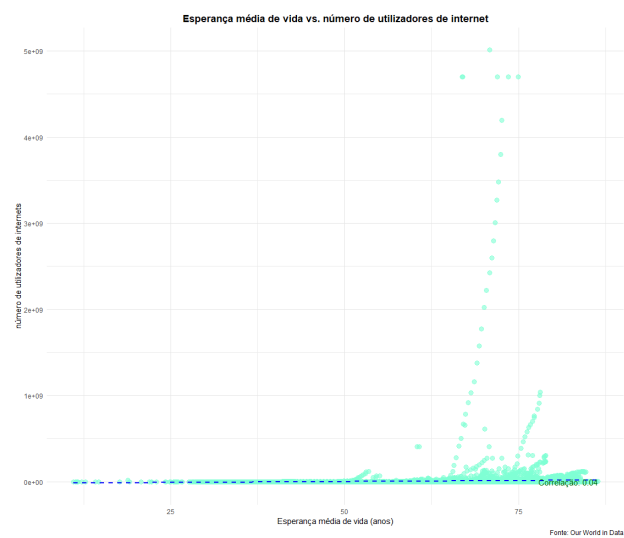
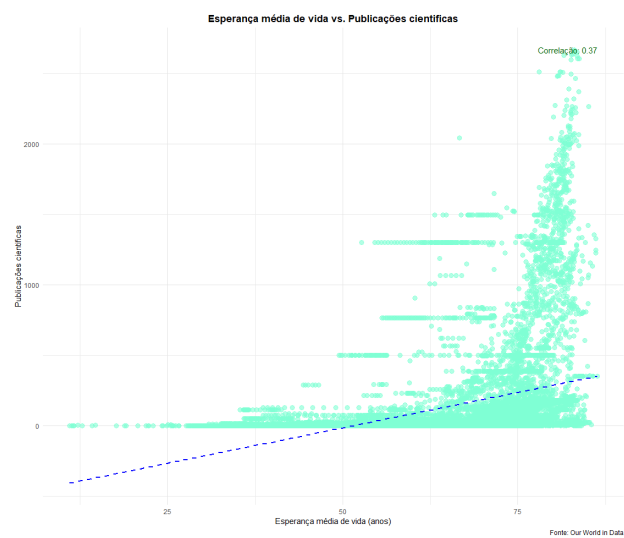
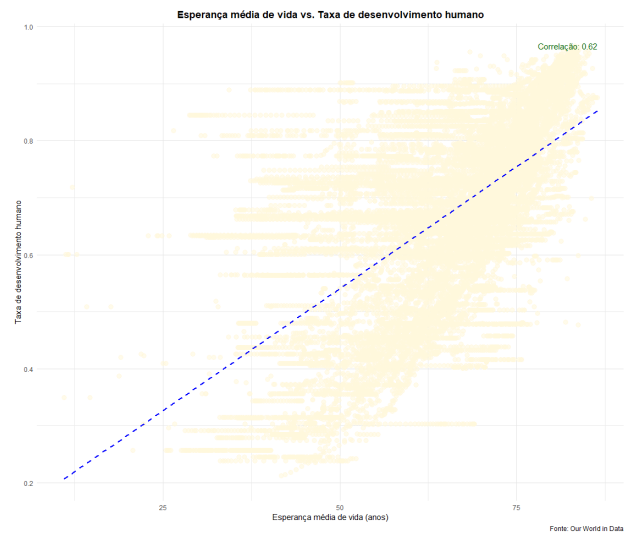
Apêndice 3 – Matriz revisão de literatura

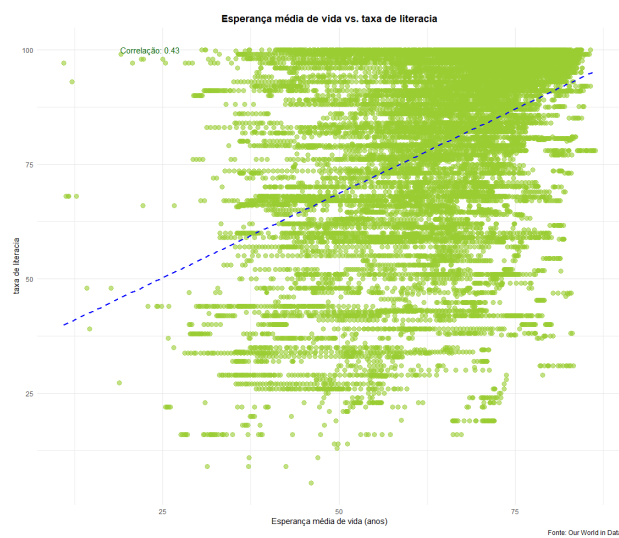
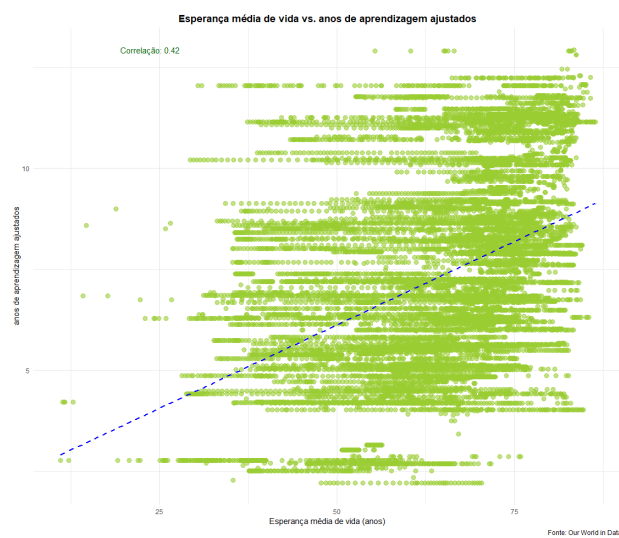
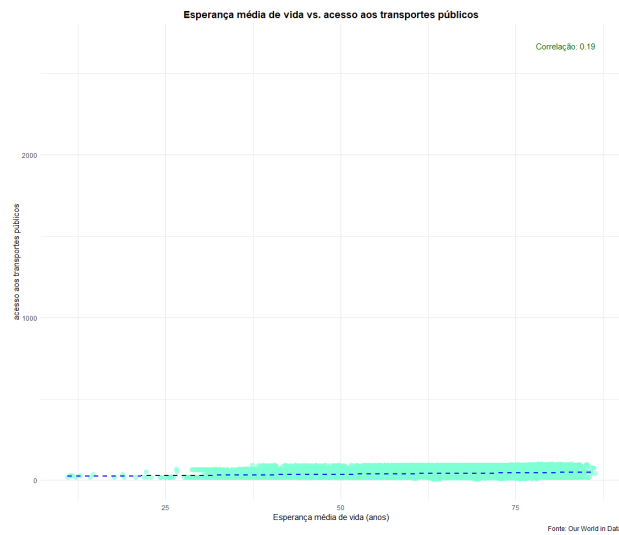
Apêndice 1

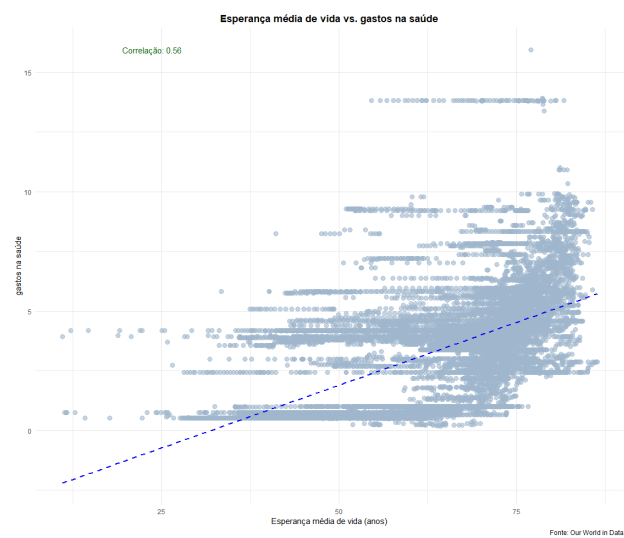
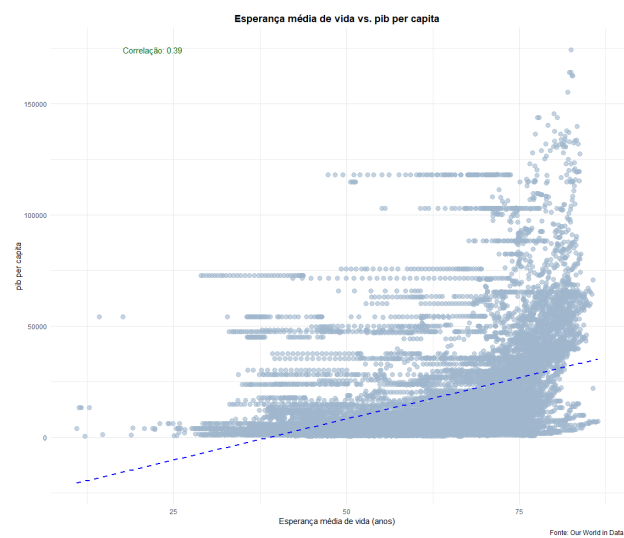
Correlação linear entre esperança média de vida e outros fatores

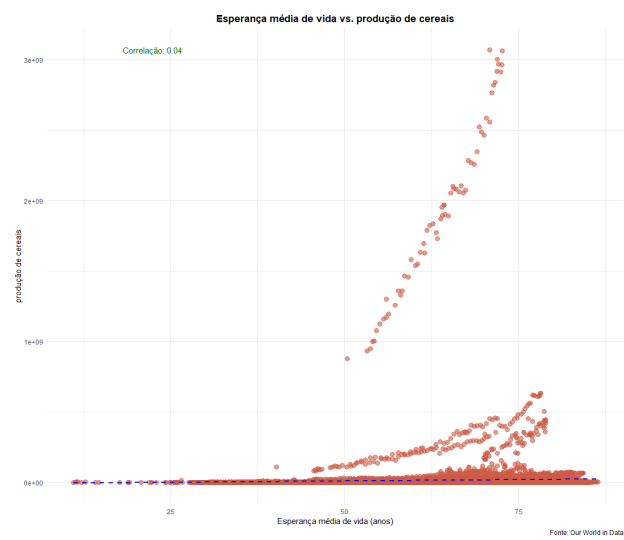
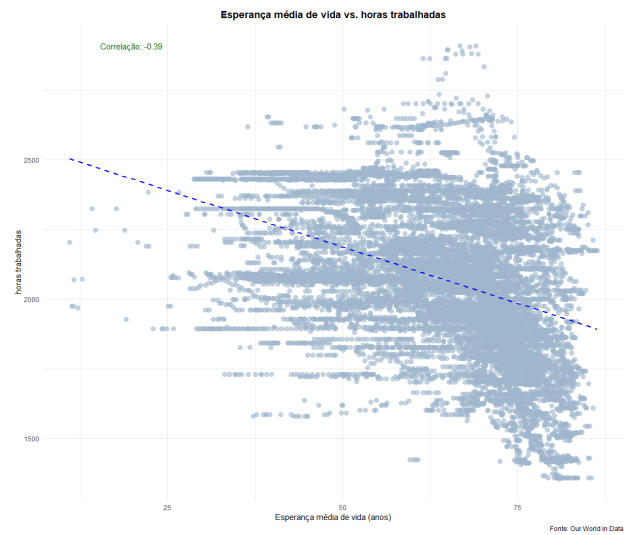


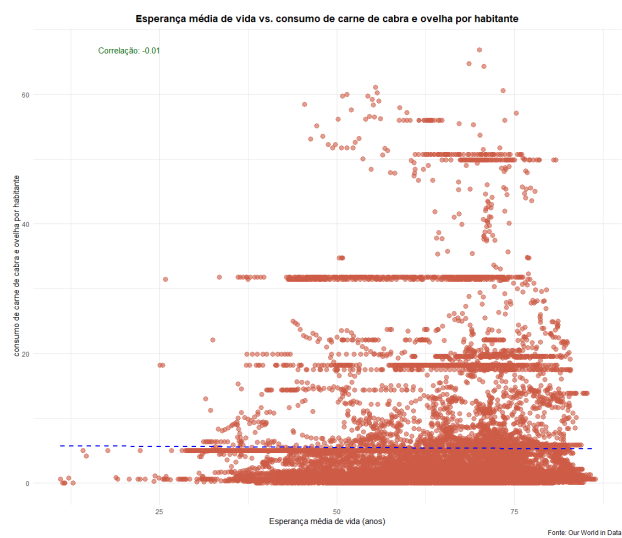
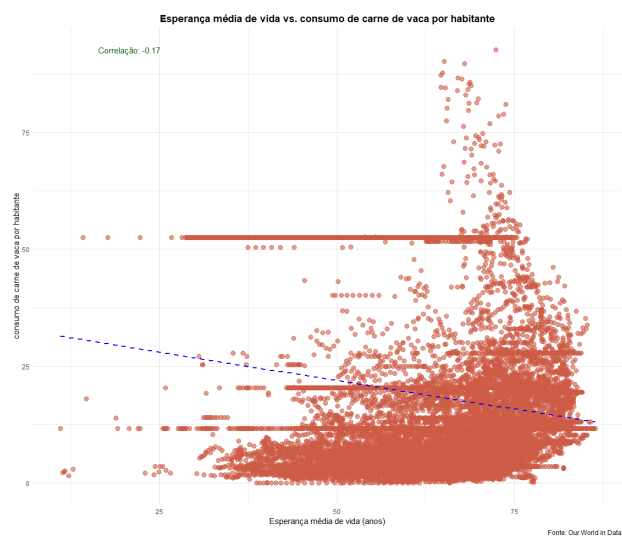
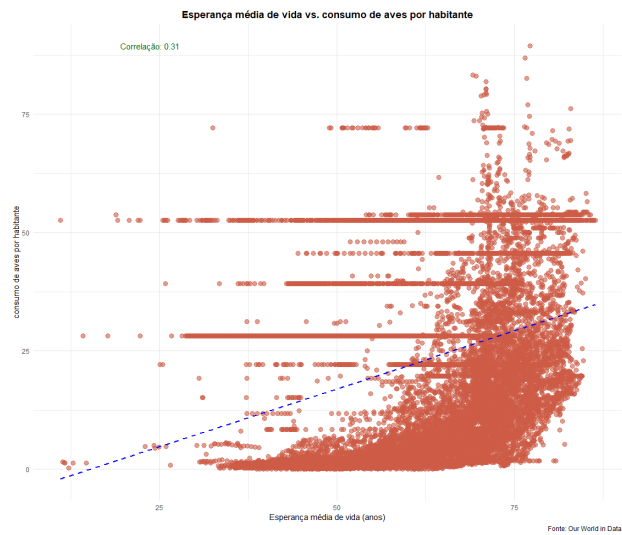


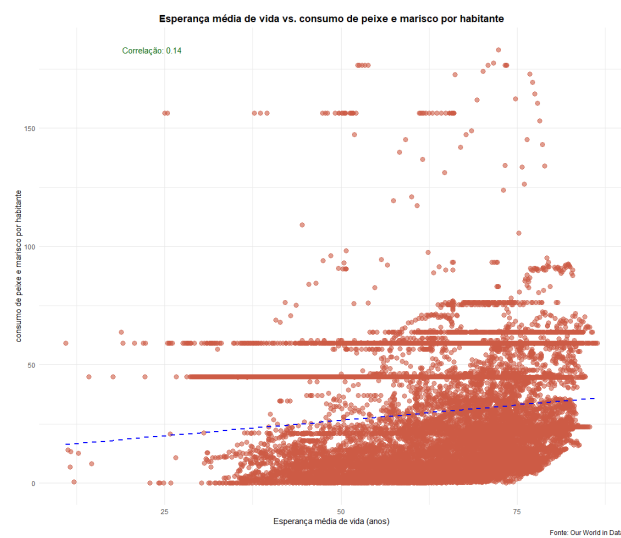
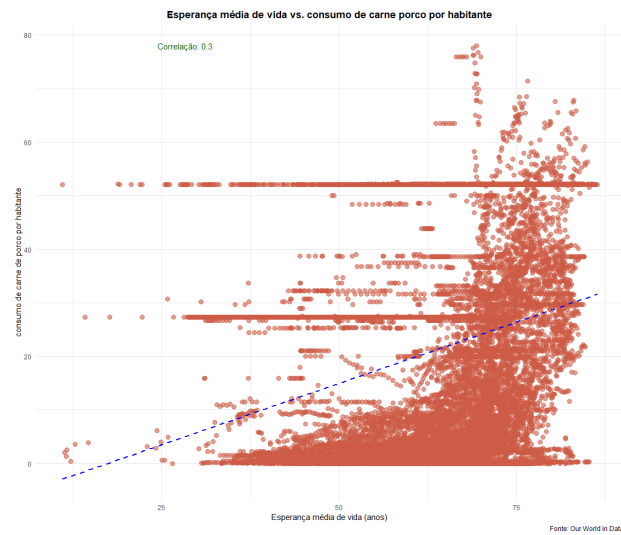


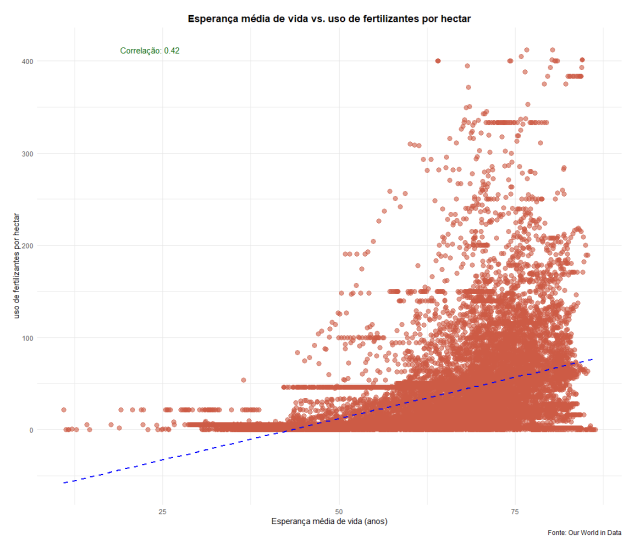
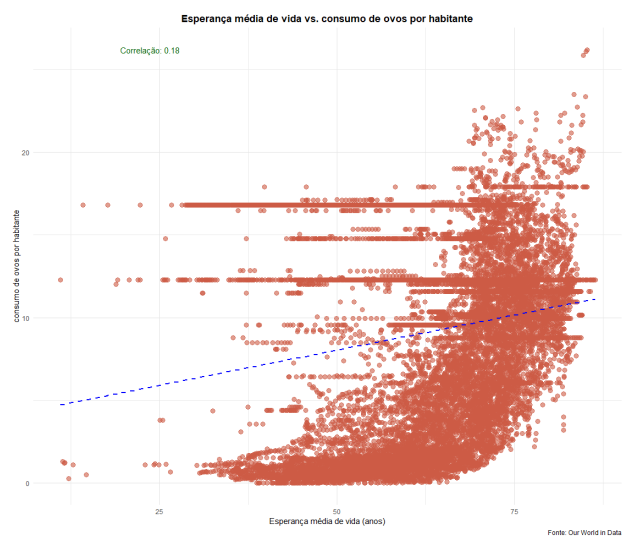
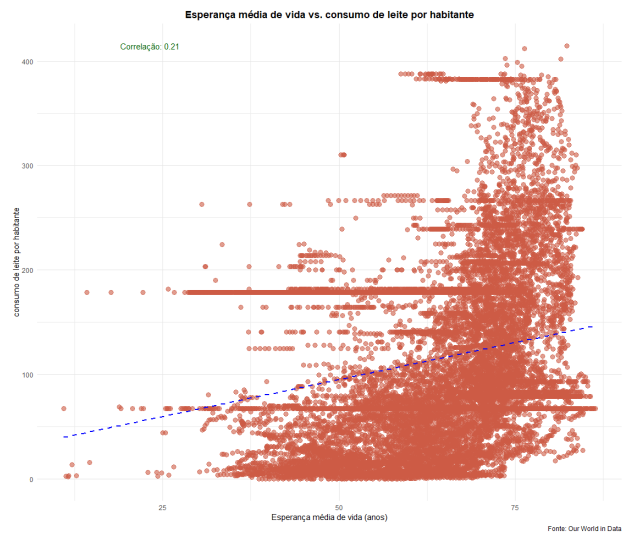


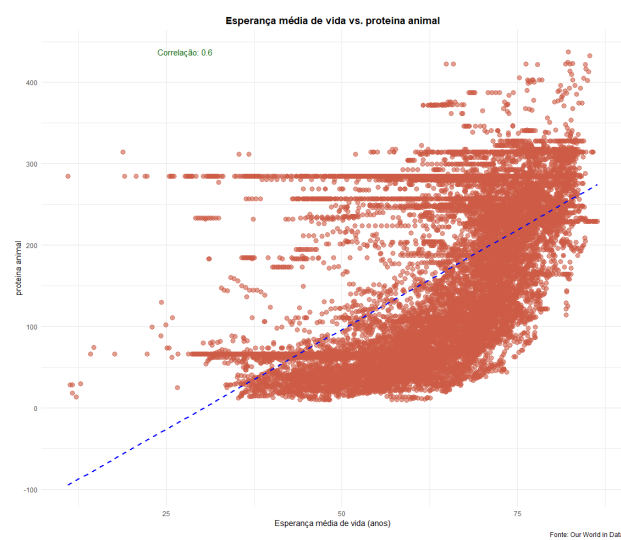
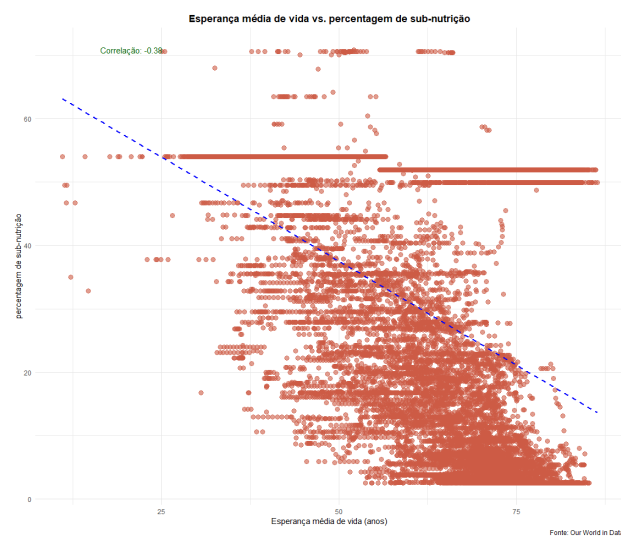
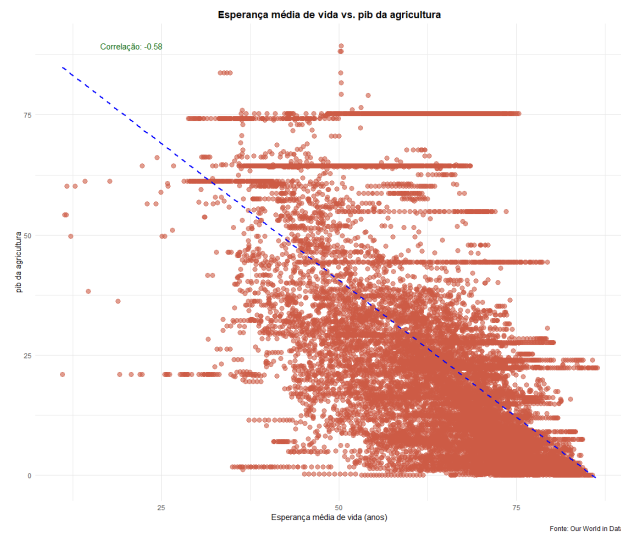


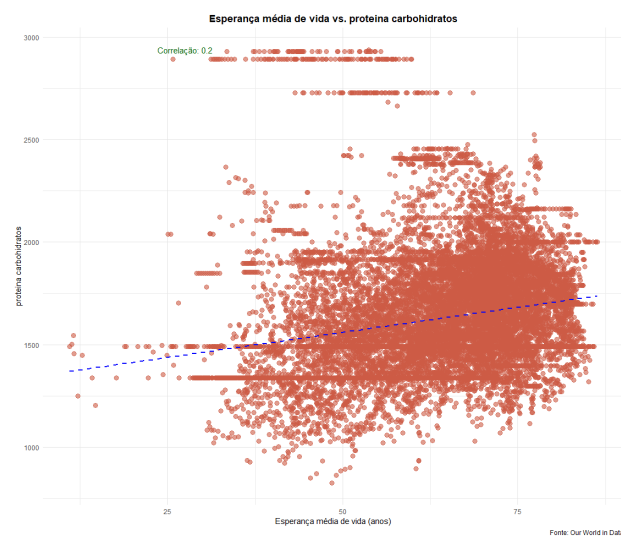
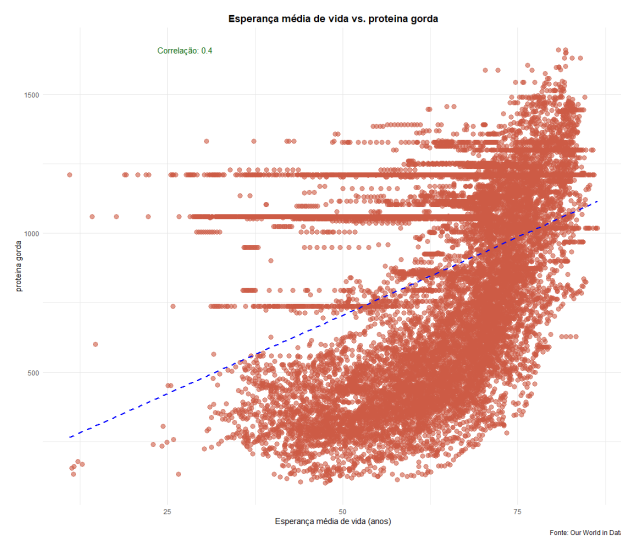
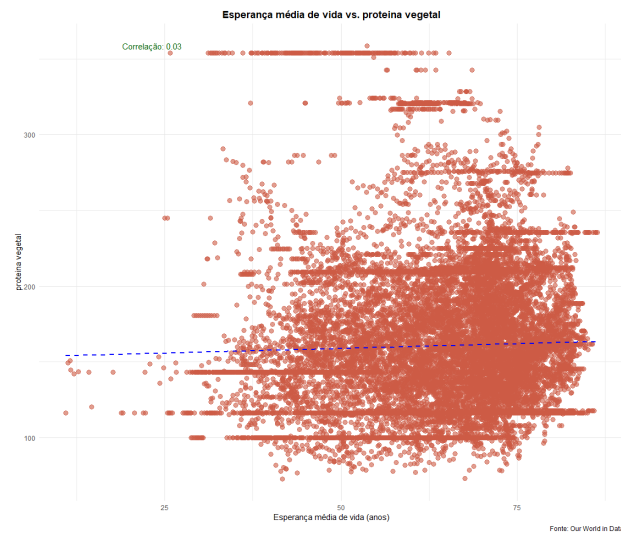


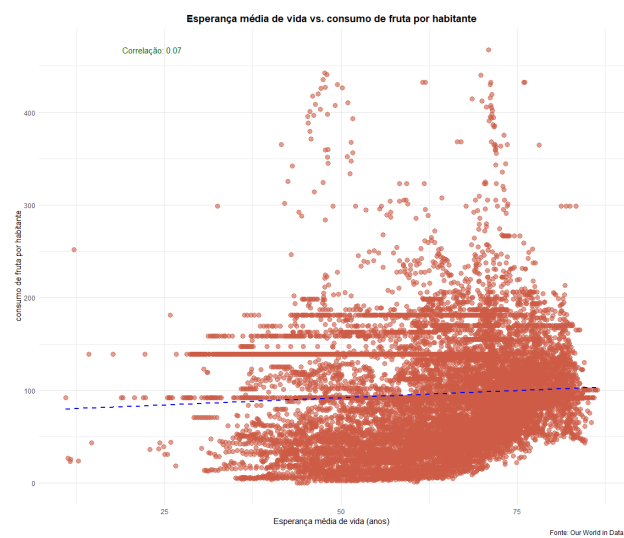
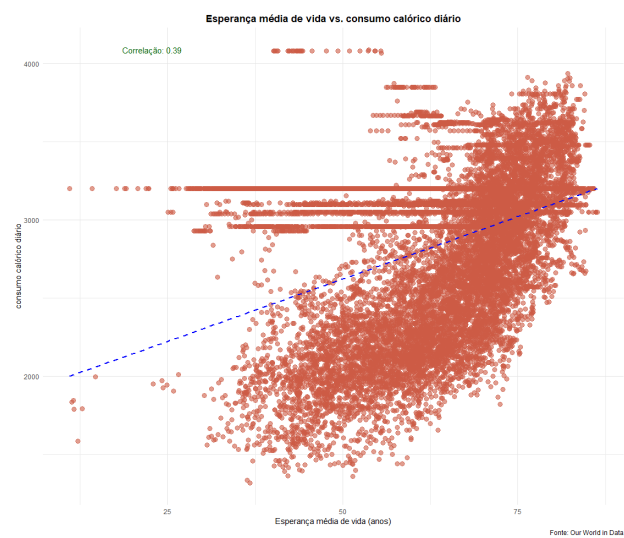
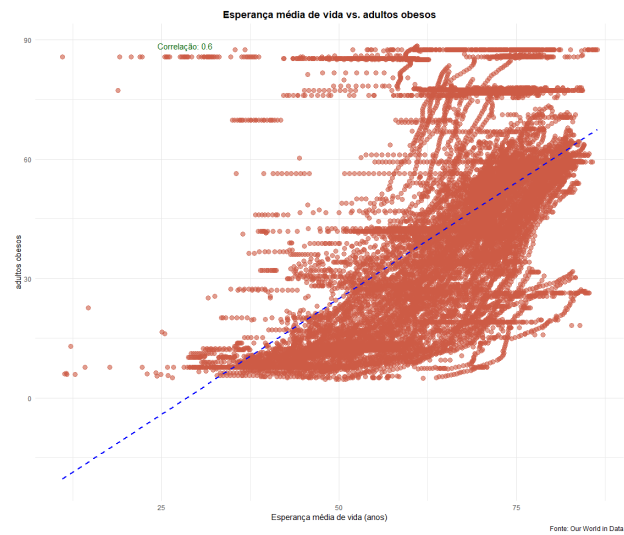


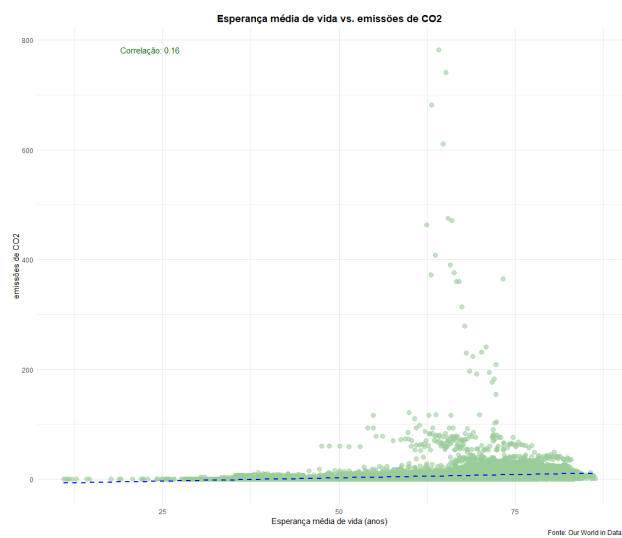
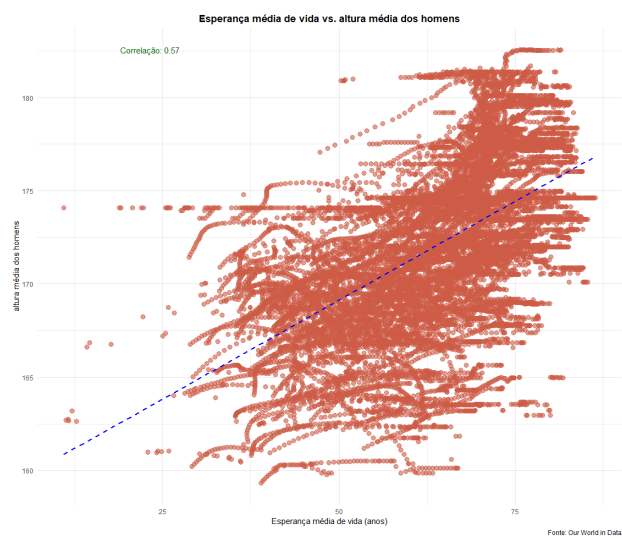
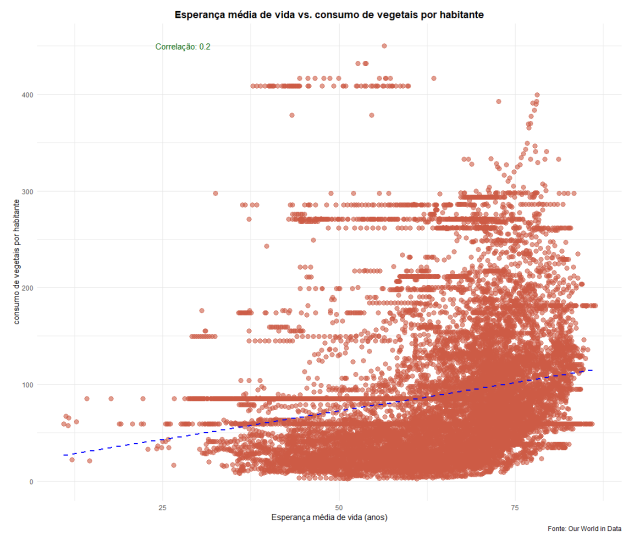


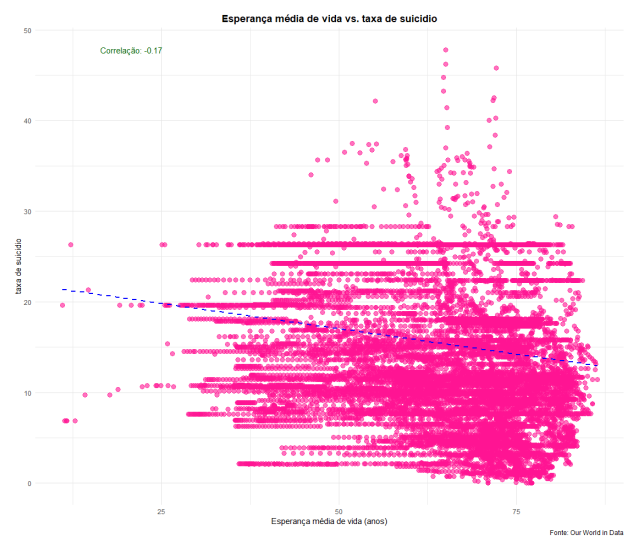
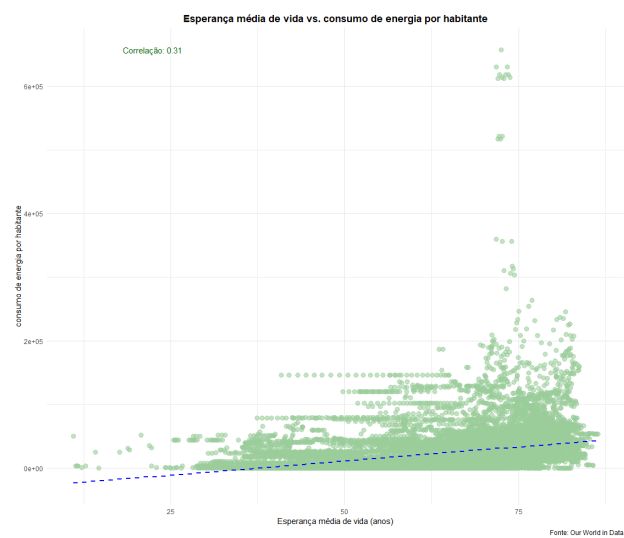
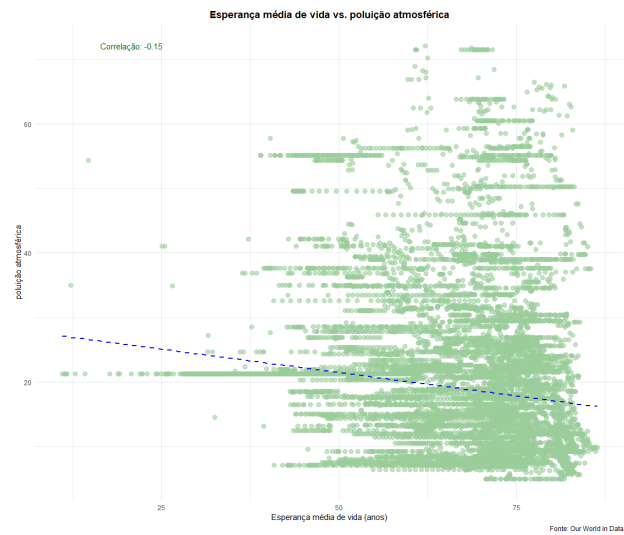


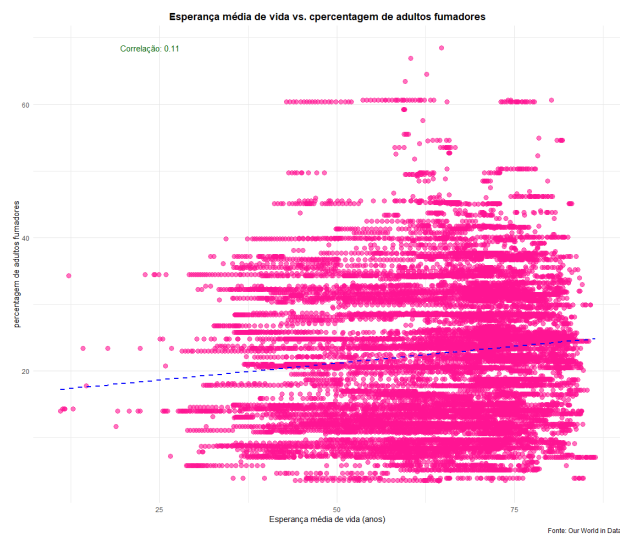
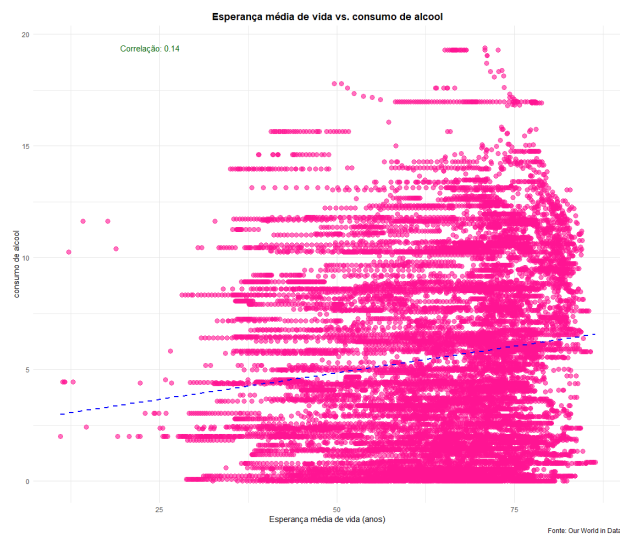


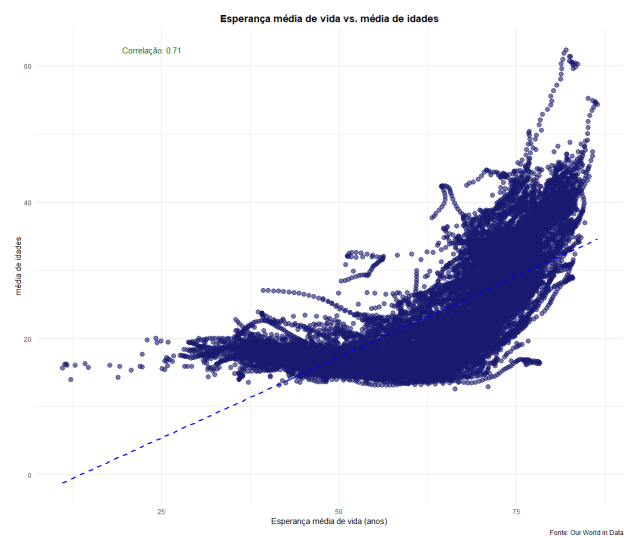
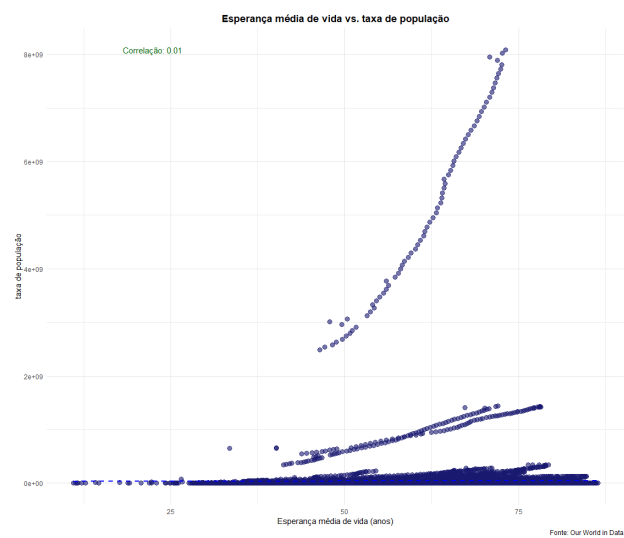


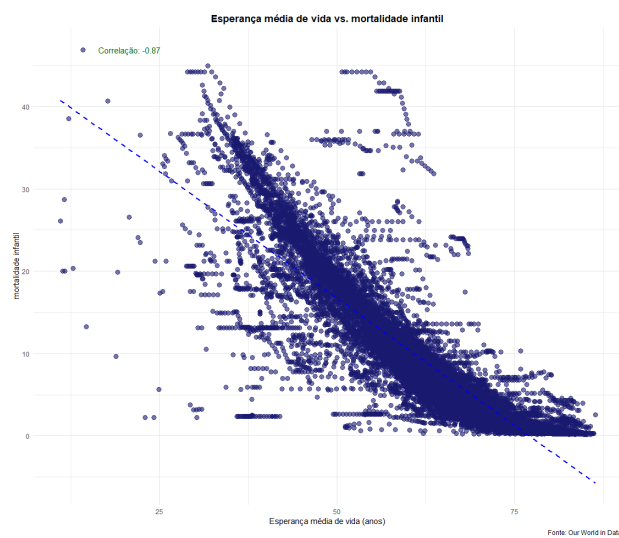
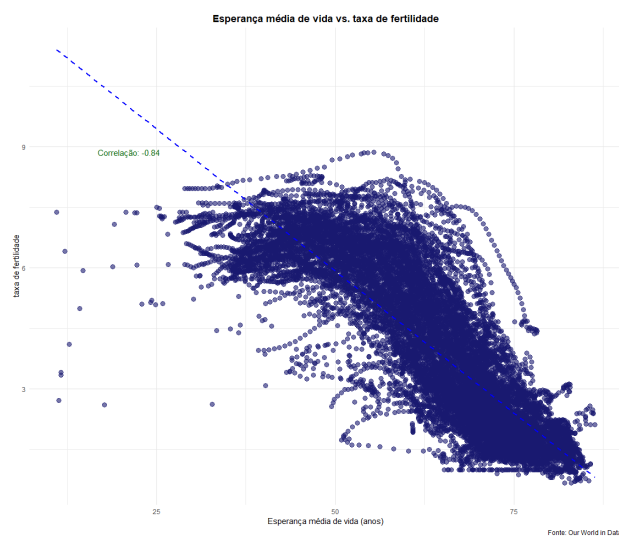
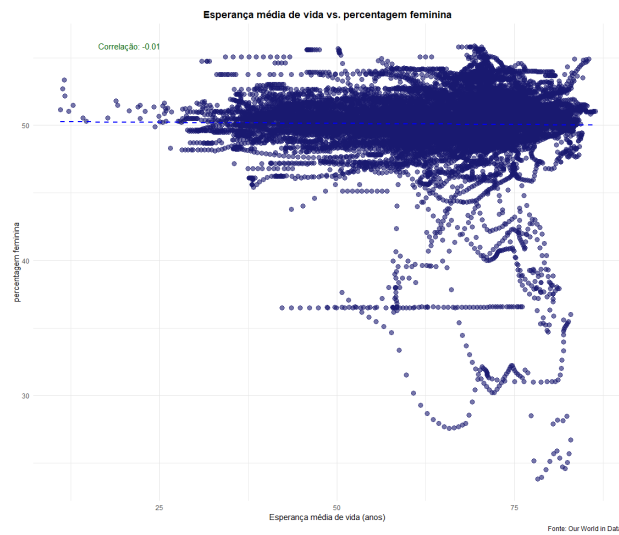


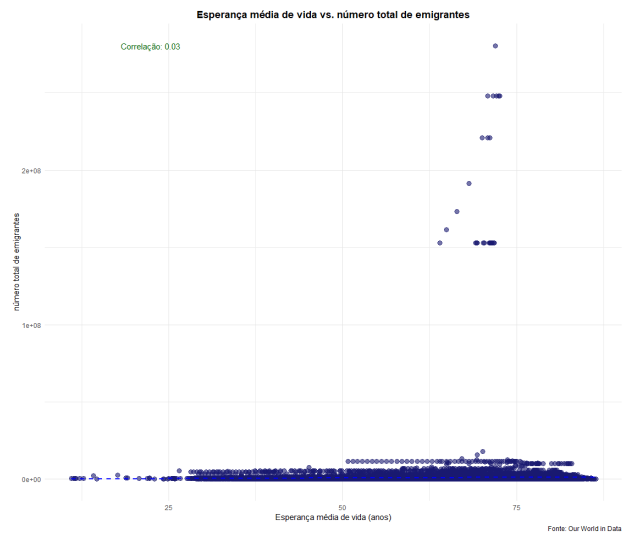






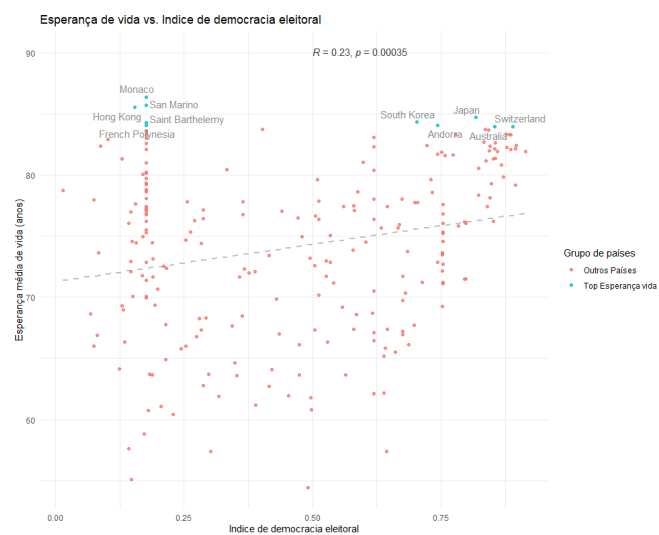
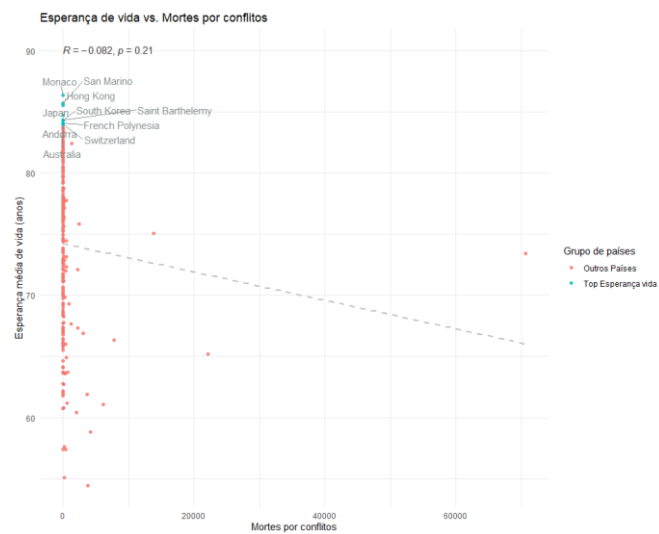


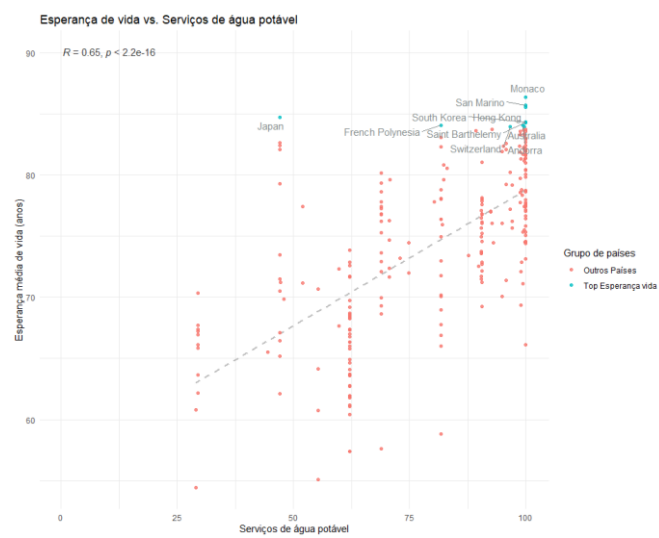
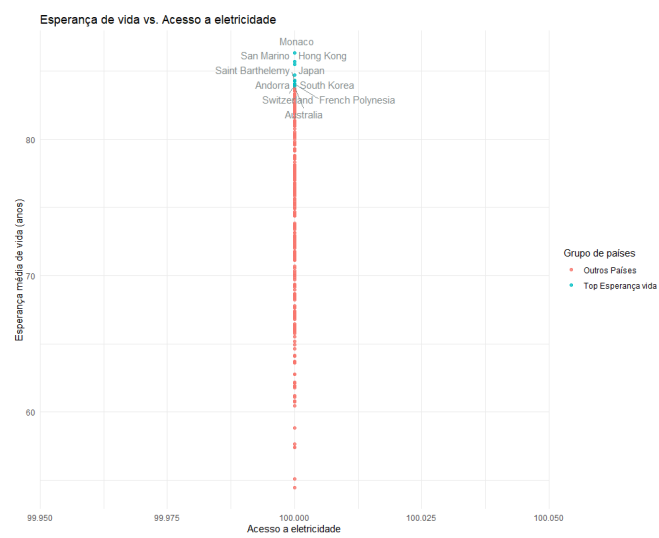
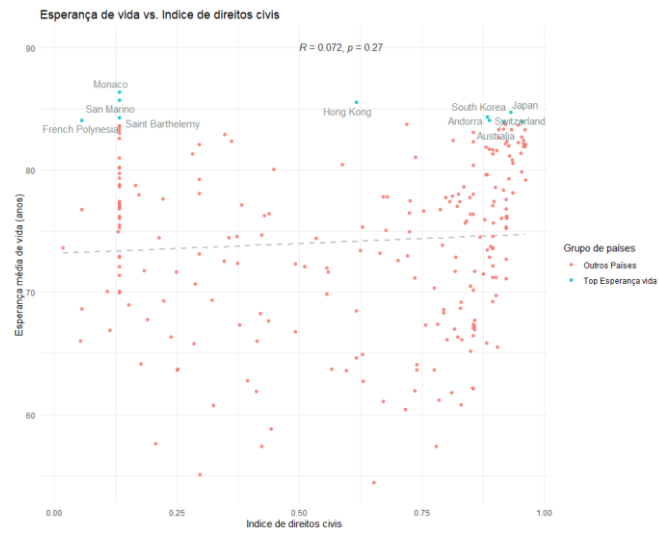


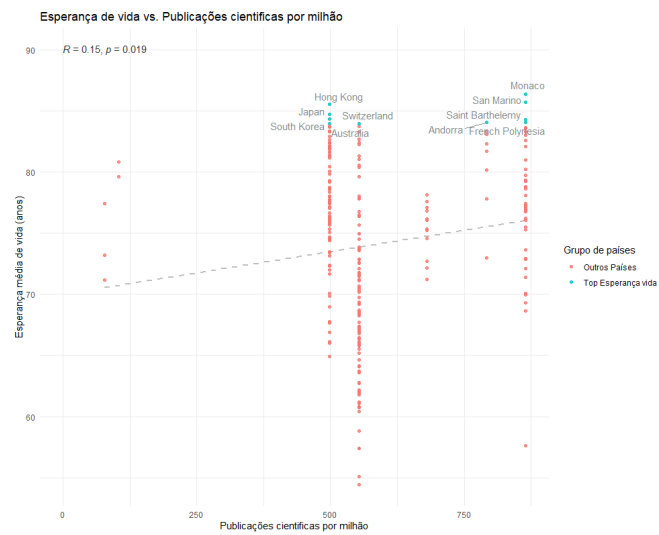
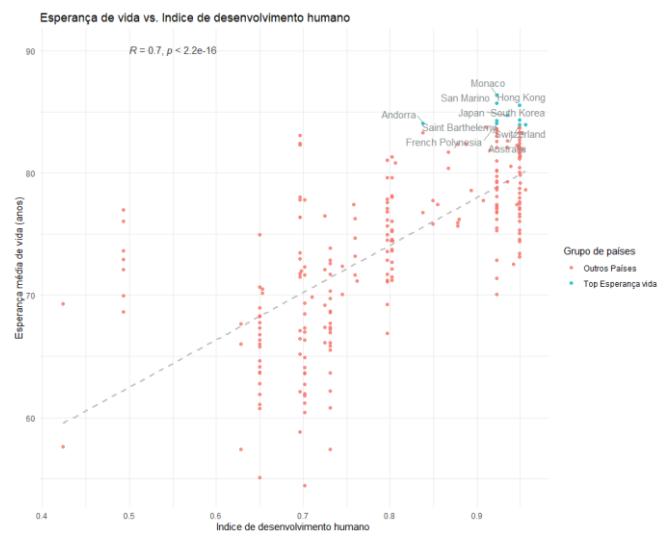
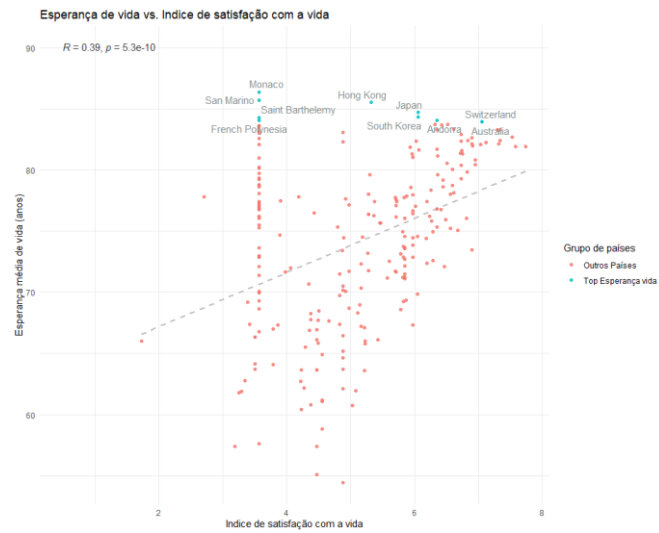


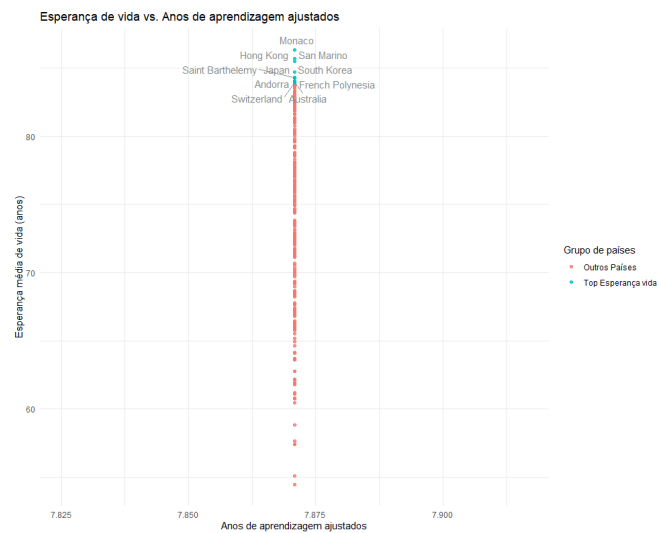
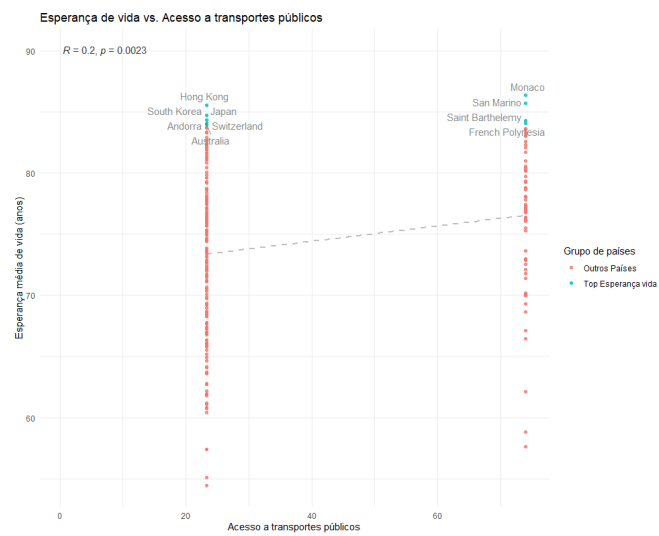
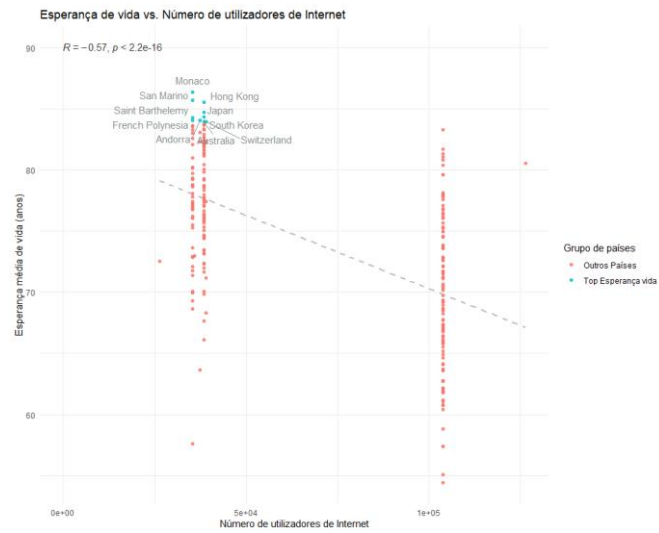
Apêndice 2

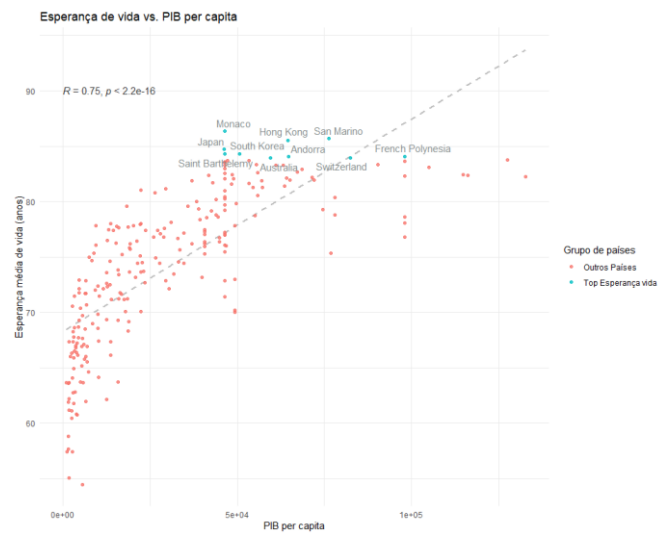
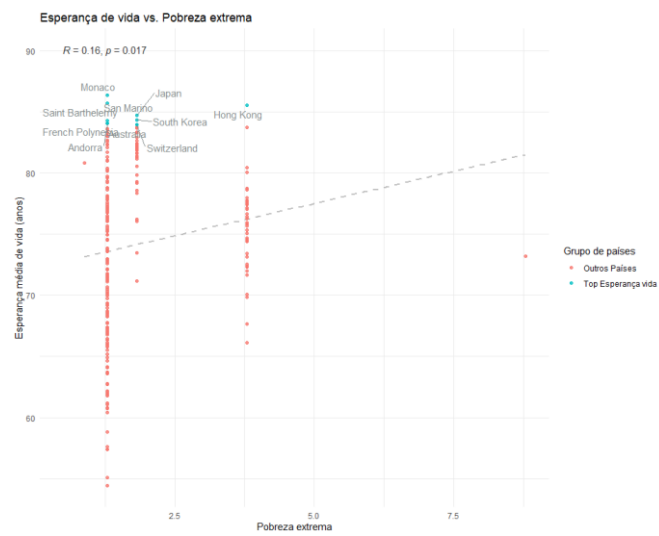
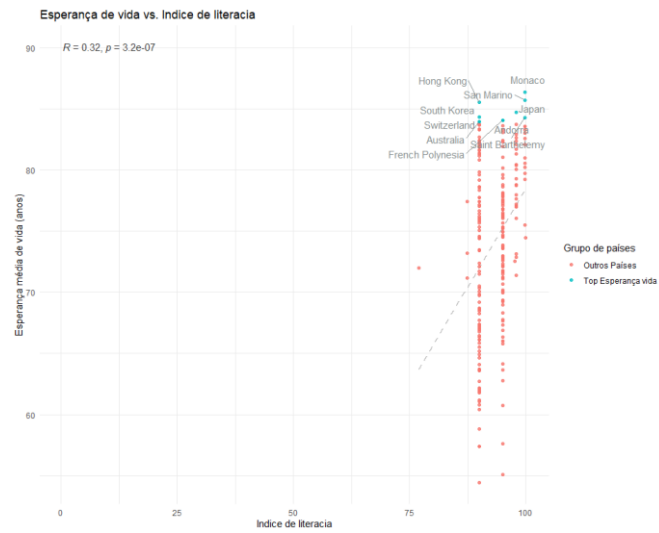
Indicadores nos países Top Esperança média de vida vs. Outros países em 2023

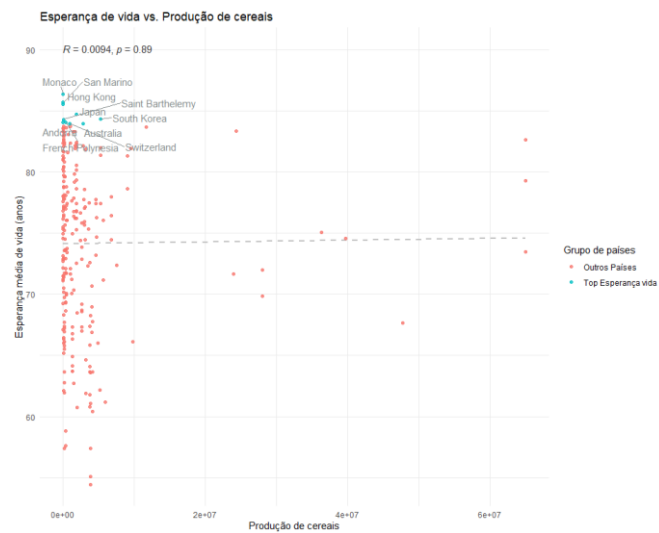
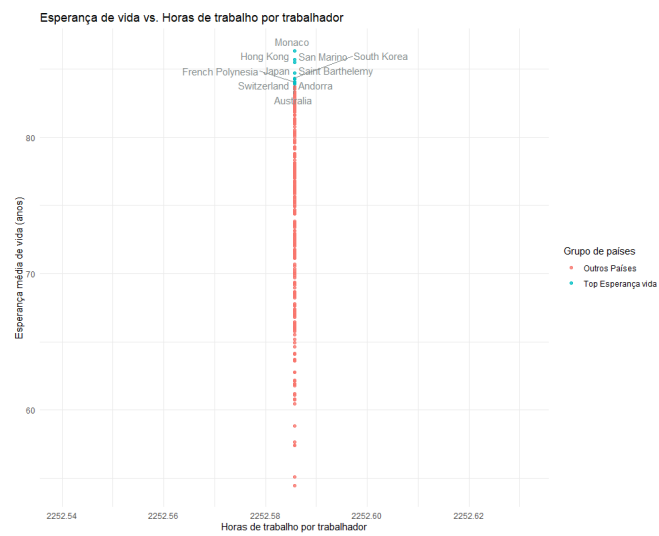
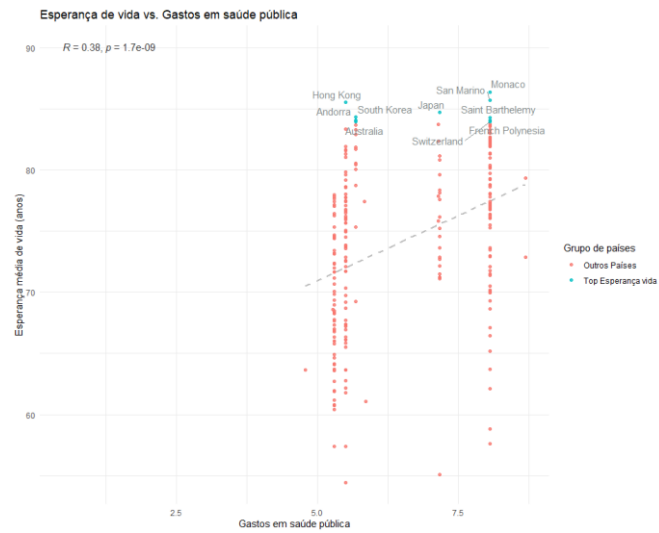


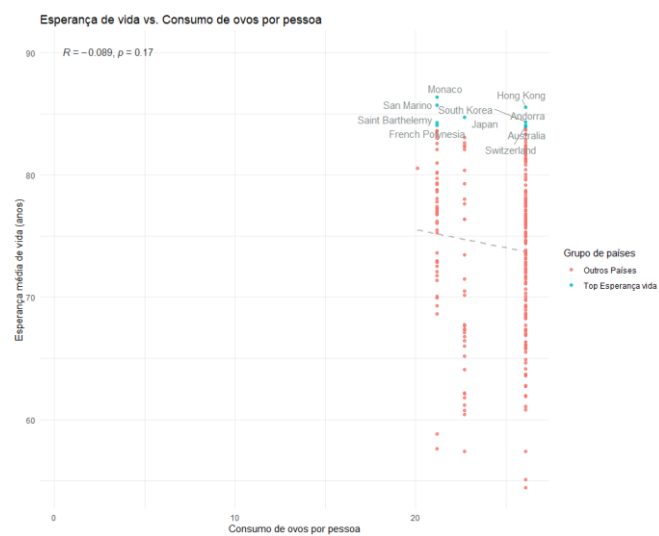
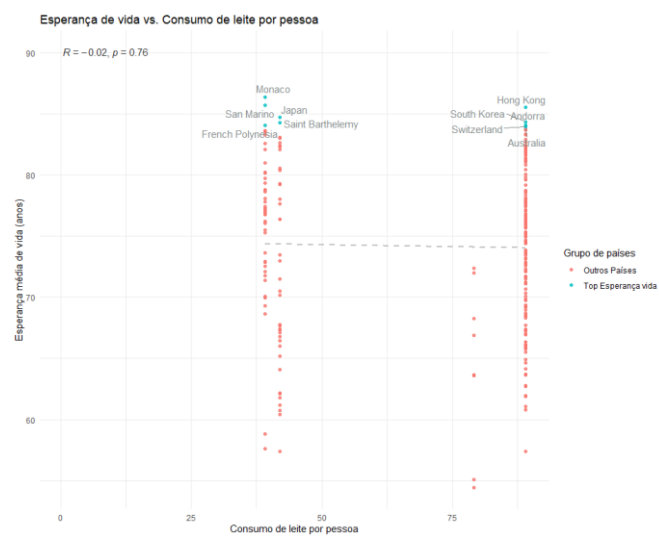
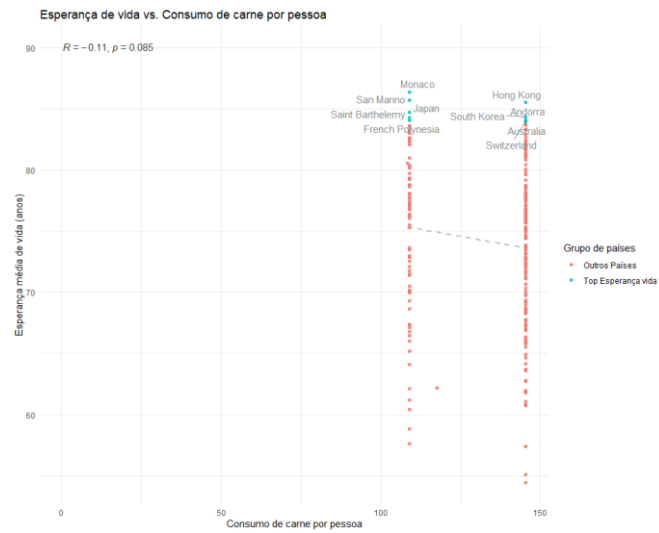


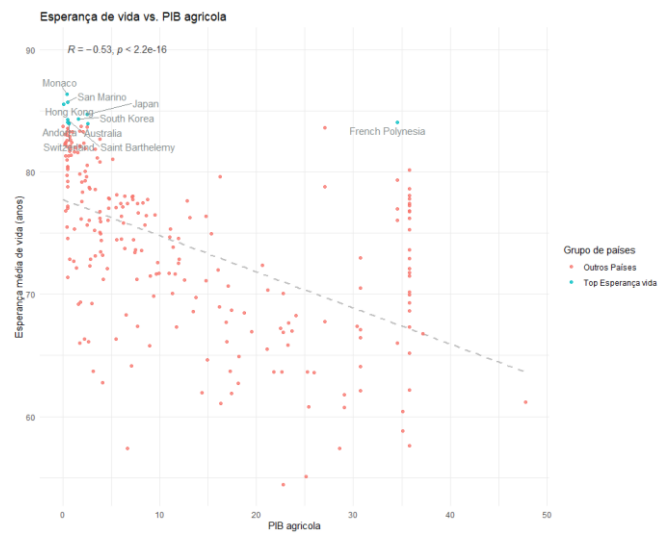
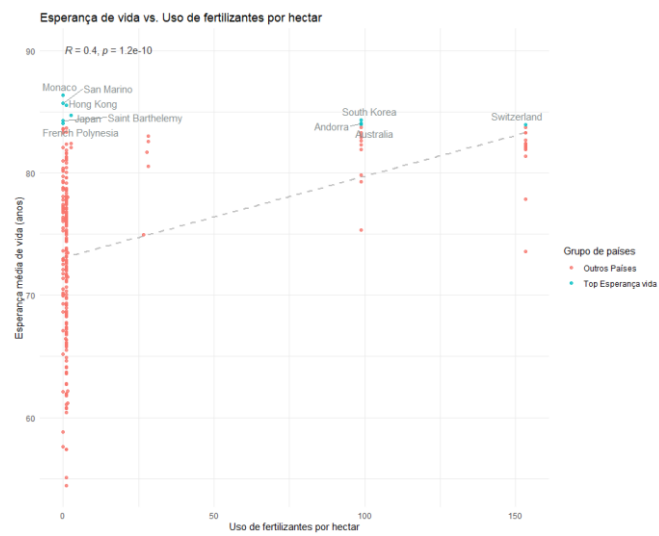
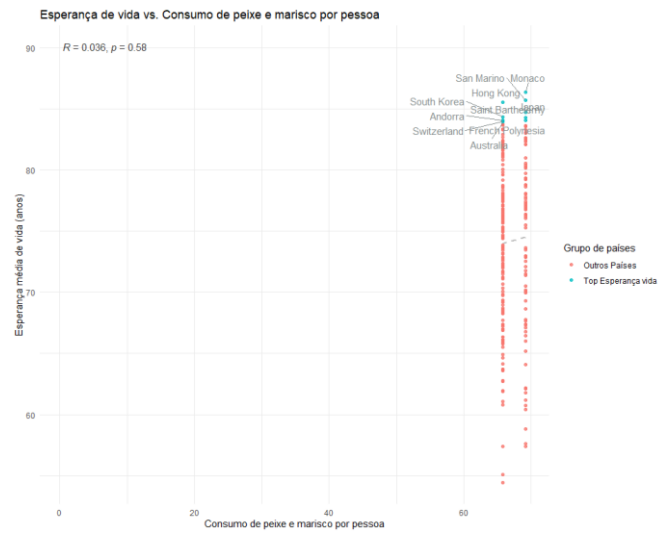


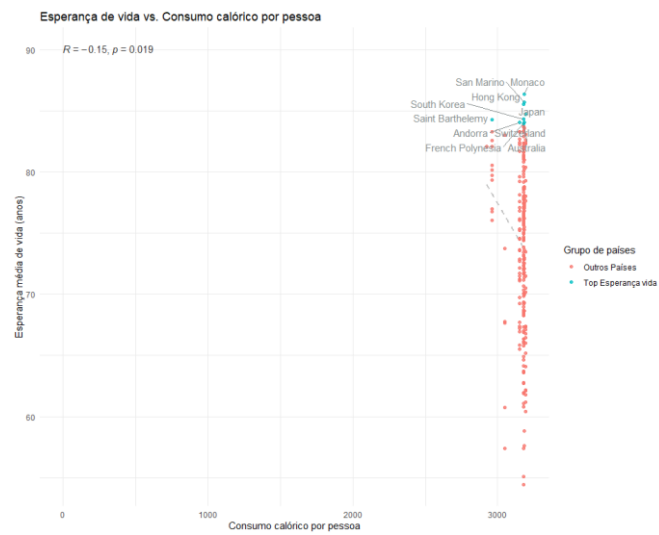
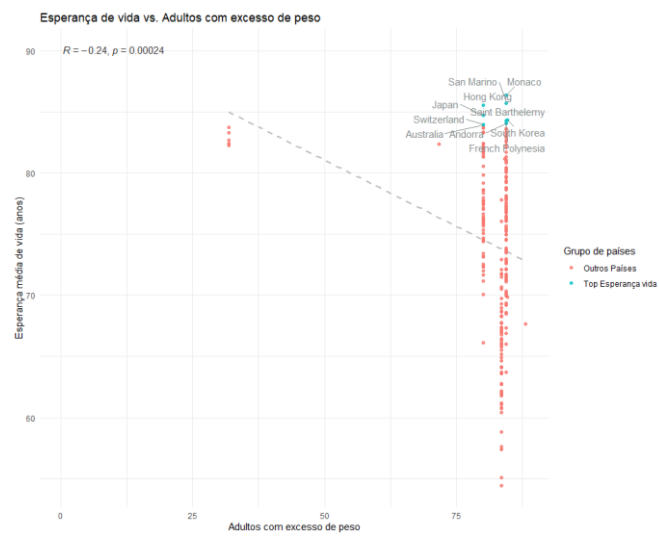
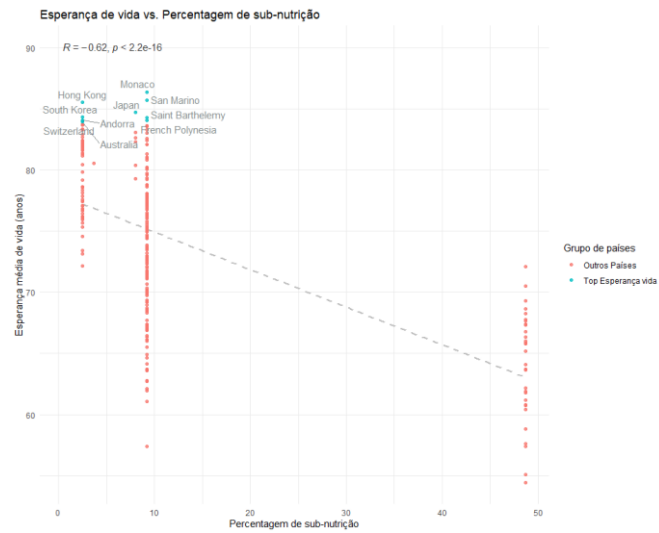


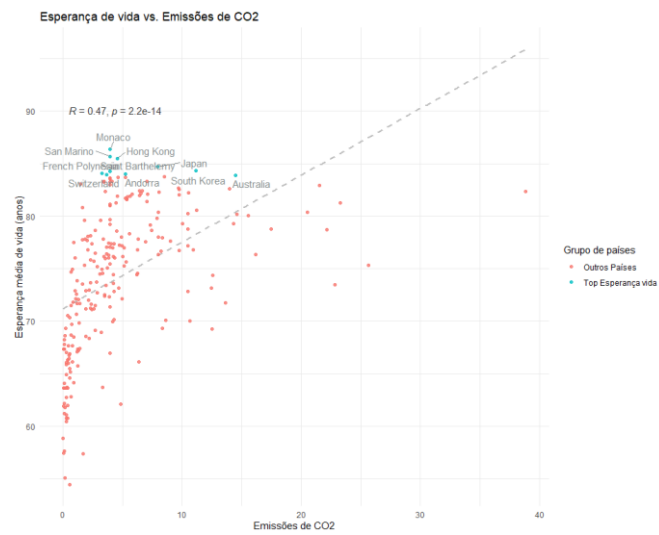
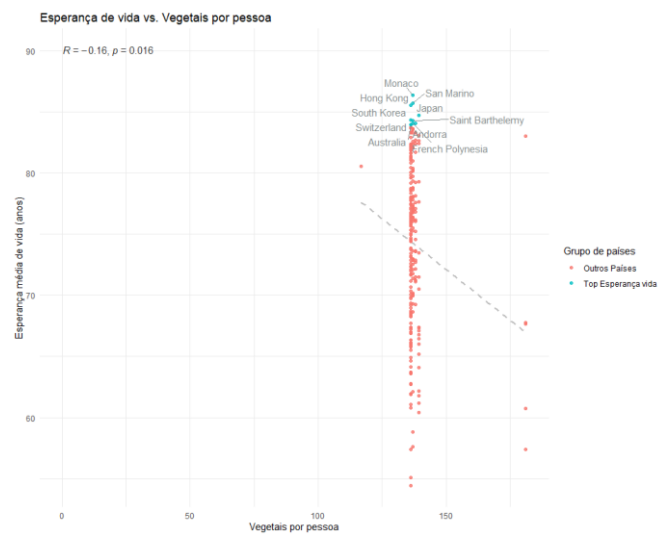
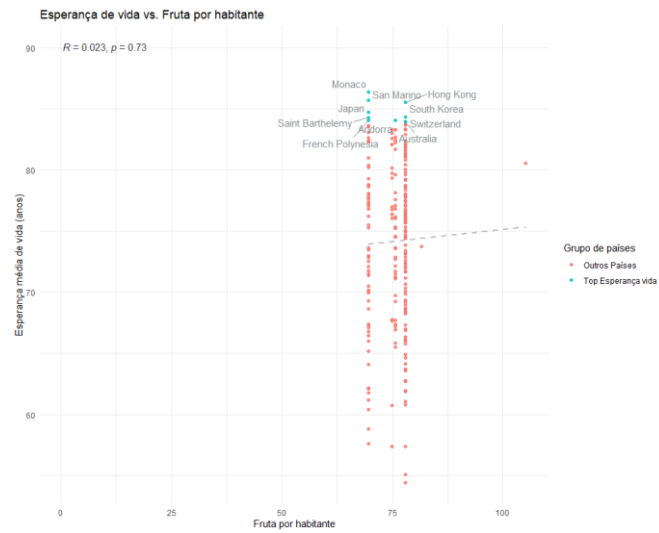


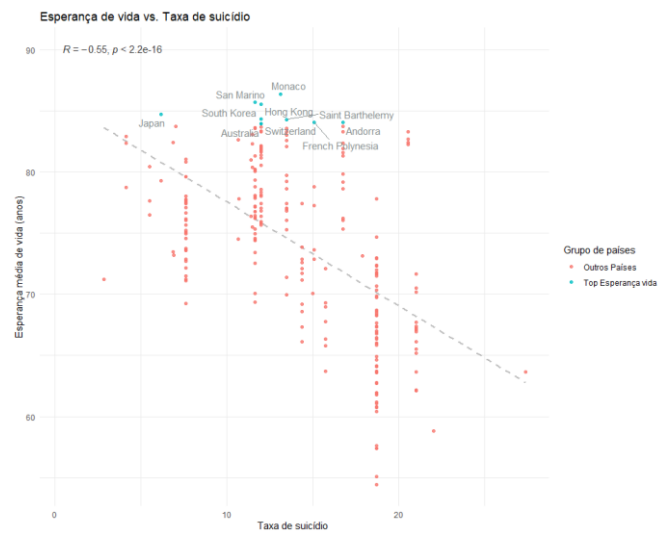
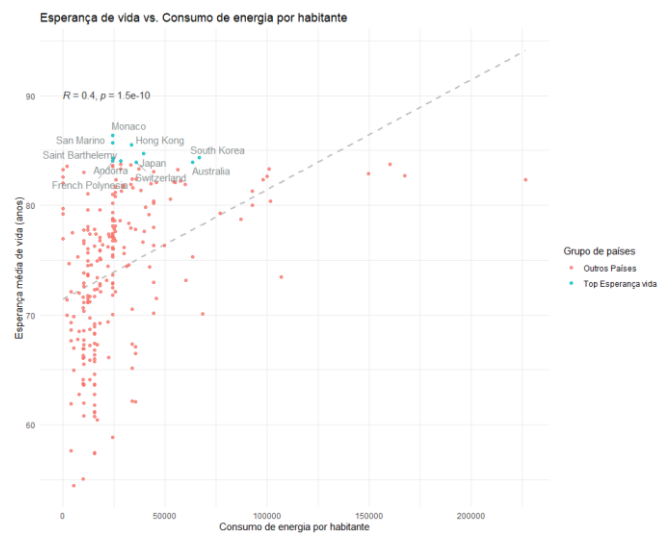
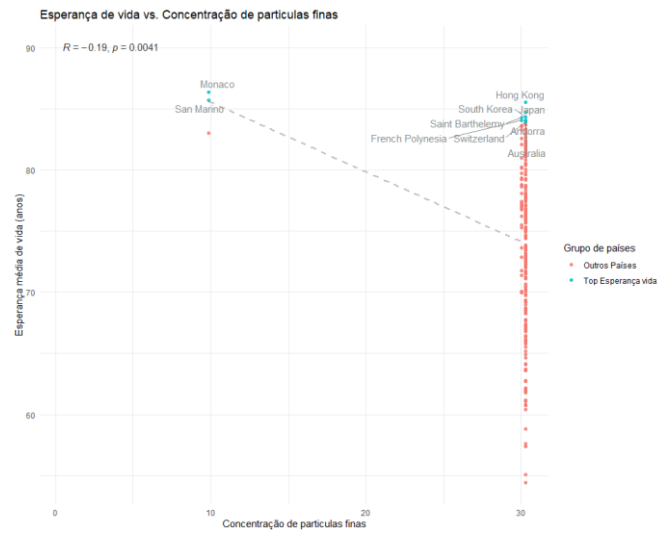


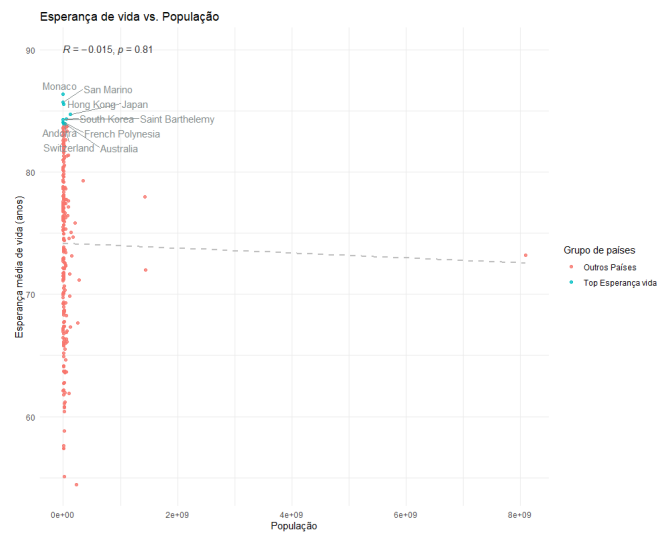
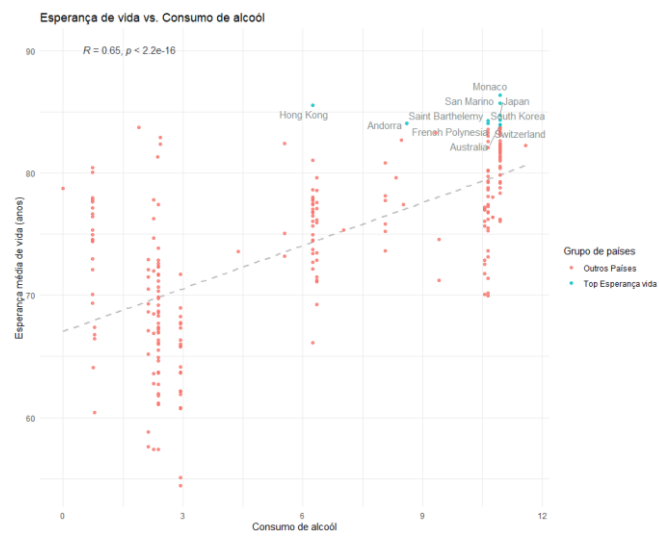
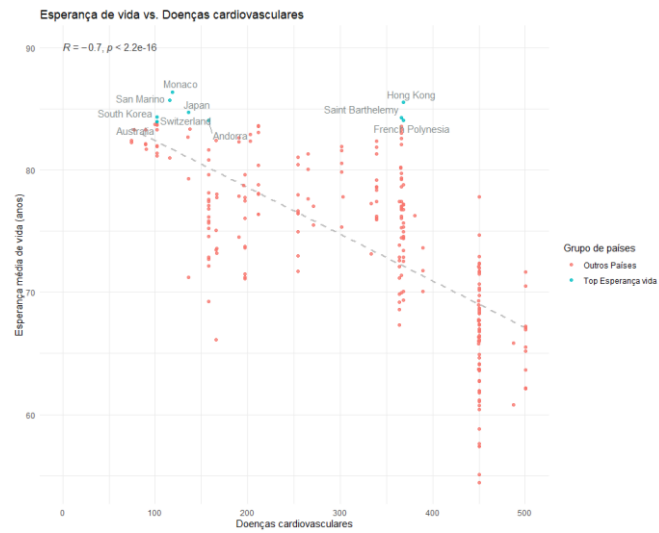


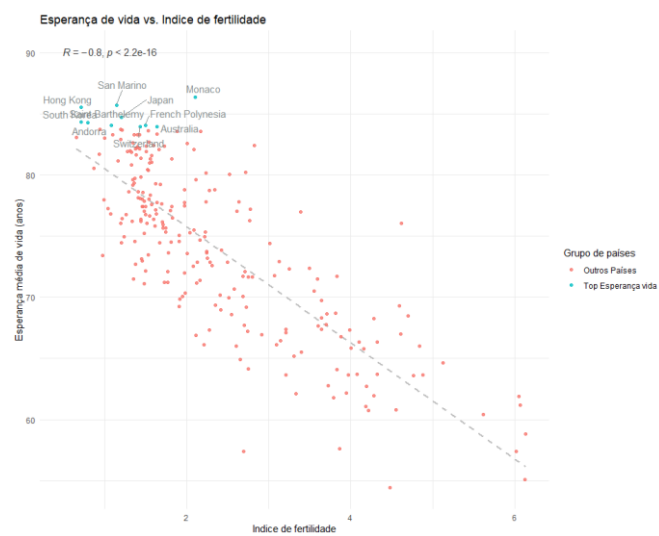
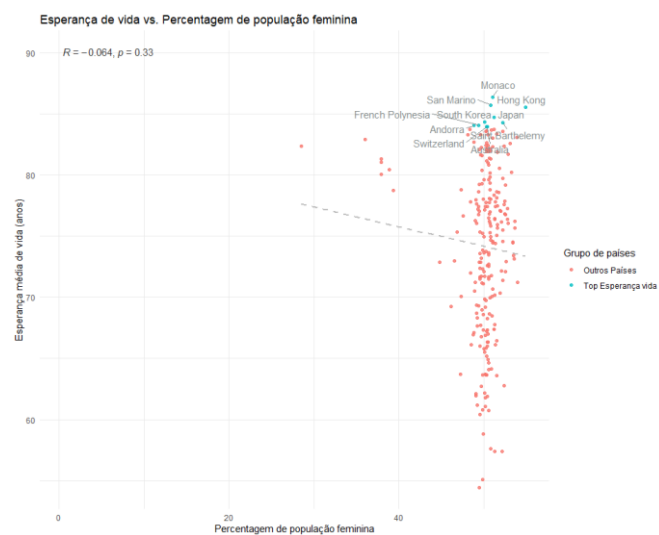
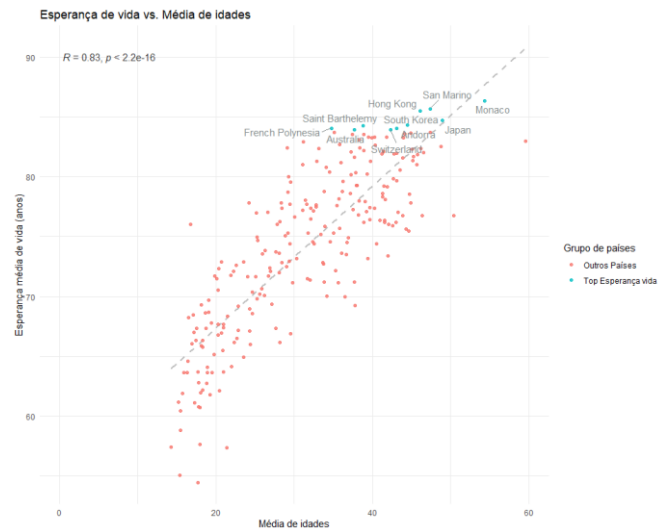


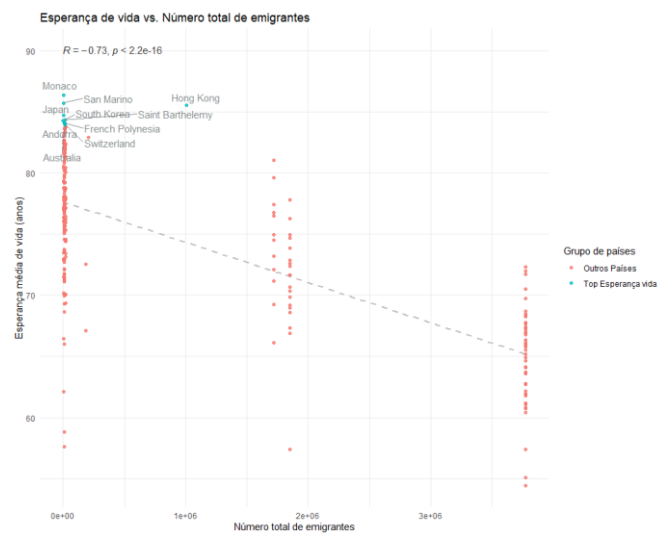
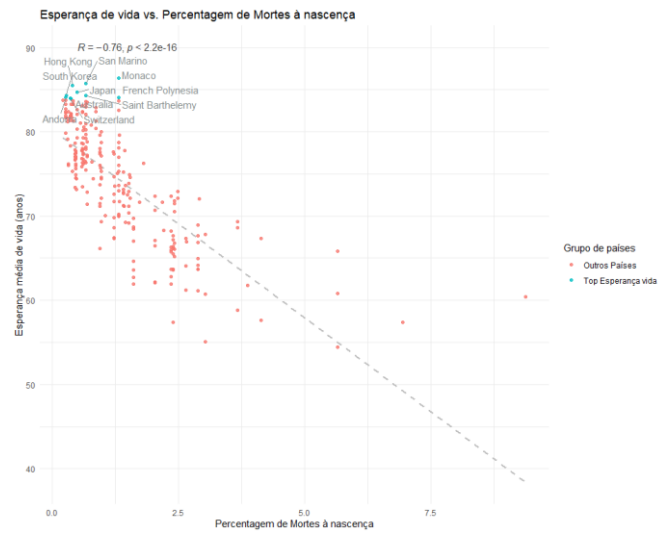












Apêndice 3

Matriz revisão da literatura

			Políticas sociais	Políticas ambientais	Políticas regionais	Saúde	Métodos de aumento	Ume fechamento ou subv	Forecast de envelhecimento da DMV	Metodologia	Fatores específicos
1	(1) Gelland, A. (2016). Global life expectancy increases by five years. <i>British Medical Journal</i> , 353. https://doi.org/10.1136/bmj.2016.090400	Rigorous journal, 24 citações									
2	(2) Mathers, C., Stevens, G., Boerma, T., White, R., & Tobias, M. (2016). Causes of international increases in older age life expectancy. <i>The Lancet</i> , 386, 540-548. https://doi.org/10.1016/S0140-6736(14)00069-9	Highly cited, 347 citações									Tabaco Doenças cardiovasculares Medicina e cuidados de saúde Dieta Exercício físico Redução de stress
3	(3) Budka, J., Chen, M., Fendrick, A., & Olfert, D. (2020). Contributions Of Public Health, Pharmaceuticals, And Other Medical Care To US Life Expectancy Changes, 1990-2015. <i>Health Affairs</i>	Observational Study, 41 citações									Tabaco Acidentes viação Doenças cardiovasculares Medicina e cuidados de saúde
4	(4) Klenk, J., Rapp, K., Büchtem, G., Keil, U., & Weiland, S. (2007). Increasing life expectancy in Germany: quantitative contributions from changes in age- and disease-specific mortality. <i>European Journal of Public Health</i> , 17, 325-328. https://doi.org/10.1093/eurpub/ckn007	Observational study, Rigorous journal, Highly cited, 85 citações									Doenças cardiovasculares Condições políticas e sociais Níveis de vida saudáveis
6	(6) Crimmins, E. (2015). Lifespan and Healthspan: Past, Present, and Promise. <i>The Gerontologist</i> , 55, 6, 803-11. https://doi.org/10.1093/geronl/gnv130	Rigorous journal, Highly cited, 448 citações									
7	(7) Bourg, É. (2012). Forecasting continuously increasing life expectancy: What implications? <i>Ageing Research Reviews</i> , 11, 325-328. https://doi.org/10.1016/j.arr.2012.01.002	23 citações									
8	(8) Vallin, J. (2010). Will life expectancy increase indefinitely by three months every year. <i>Population and societies</i> .	25 citações									Doenças cardiovasculares (antes, doenças infecciosas)
9	(9) Eggleston, K., & Fuchs, V. (2012). The New Demographic Transition: Most Gains in Life Expectancy Now Realized Late in Life. <i>Health & the Economy Journal</i> . https://doi.org/10.1257/HEP.26.3.1	Highly cited, 129 citações									
10	(10) Oeppen, J., & Vaupel, J. (2002). Broken Limits to Life Expectancy. <i>Science</i> , 296, 1029-1031. https://doi.org/10.1126/SCIENCE.1069675	Highly cited, 2482 citações									
11	(11) Mackenbach, J. P., Rubio Valverde, J., Bopp, M., Brønnum-Hansen, H., Deboosere, P., Kalediene, R., Kovács, K., Leinsalu, M., Martikainen, P., Menietti, G., Regidor, E., & Nusselder, W. J. (2016). Determinants of life expectancy: A panel data approach. <i>Asian Economic and Financial Review</i> , 6(11), 1263-1267										Tabagismo Baixos rendimentos Excesso de peso
12	(12) Abdalrhman, M., & Mehjandi, A. S. (2015). Determinants of life expectancy: A panel data approach. <i>Asian Economic and Financial Review</i> , 6(11), 1263-1267										
13	(13) Raleigh, V. S. (2016). Trends in life expectancy. <i>OECD Health Working Papers No. 108</i> . https://dx.doi.org/10.1767/223156b_en										
14	(14) Yan, F., Agha, L., Sadeghi, H., Faraji, D., & S. (2022). The socioeconomic and political determinants of life expectancy in selected countries of the world. <i>Soc Determinants Health</i> , 2022(81), 1-8. DOI: 10.1016/j.socdet.2022.08.001										PIB per capita Taxa de alfabetização Gastos com saúde Estabilidade política
15	(15) Fadnes, L. T., et al. (2023). Life expectancy can increase by up to 10 years following sustained shifts towards healthier diets in the UK. <i>Nature Food</i> . https://doi.org/10.1038/s43016-023-0086-6										
16	(16) Kim, I. H. (2010). Age and cause specific contributions to increase in life expectancy at birth in Korea										
17	(17) Sugiyama, Y. (2010). Rapid increase in Japanese life expectancy after World War II	31 citações									Reformas da terra Trabalho comunitário Ingestão de energia e proteína Educação
18	(18) Yan, Zheng, Qingrong Chang, Paul Siu, Kai Yip. Understanding the Increase in Life Expectancy in Hong Kong: Contributions of Changes in Age and Cause-Specific Mortality										Reformas da saúde Redução de doenças crônicas Acesso a serviços de saúde Impacto negativo do pneumonia
19	(19) Mehrez Abbasi-Kangvari, R. Malekpour, Masoud Masinadi, Saeed Moghaddam, S.H. Shamsi, Ne Rezaii, Na Rezaii, Bagher Larjani, Farshad Farzadfar, Ali H. Mokdad, Mehrez Abbasi-Kangvari	Scholar, 22 citações									Polição do ar é devida a particular em
20	(20) Raj Chetty, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, David Olfert. The Association Between Income and Life Expectancy in the United States										Rendas e desigualdade Tabagismo Obesidade Exercício físico Polição do ar Acesso a cuidados de saúde
21	(21) Jacek A Kopeck, Eric C Sayre, Benajir Shams, Linda CL, Hui Xie, Lynne M Feehan, John M Estabrook (2022). The Impact of 55 Risk Factors on Life Expectancy in Canada: Findings from a New Risk Factor Model										Tabagismo Pressão arterial Obesidade Colesterol Polição
22	(22) Wei-Cheng Lin, Tsung-Hao Ho, Cheng-Yu Shih, Hsien-Ho Lin, Jing-Shiang Hwang (2024) Impact of Healthy Lifestyle Factors on Life Expectancy and Lifetime Health Care Expenditure: Nationwide Longitudinal Study	Scholar, 5 citações									Tabagismo Alcool em excesso Exercício físico Consumo de frutas e vegetais Acesso a cuidados de saúde
23	(23) Hao, Y., Dupre, M. E., Qi, D., Wu, B., & Zeng, Y. (2020). Access to healthcare and life expectancy among older adults in China: A longitudinal study. <i>BMC Geriatrics</i> , 20(129). https://doi.org/10.1186/s12916-020-01568-9										Condições socio-econômicas Educação Nutrição
24	(24) Reimers, C. D., Knapp, G., & Reimers, A. K. (2012). Does physical activity increase life expectancy? A review of the literature. <i>Journal of Aging Research</i> , 2012, Article 249968. https://doi.org/10.1155/2012/249968										Exercício físico

citado