



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
INSTITUTO METRÓPOLE DIGITAL
BACHARELADO EM TECNOLOGIA DA INFORMAÇÃO
IMD1153 - ANÁLISE VISUAL DE DADOS

Ian Mendes Fernandes Do Nascimento
Marcone Rafael Do Nascimento Júnior
Wellder Bernardo De Oliveira

RELATÓRIO REFERENTE A VISUALIZAÇÕES E ANÁLISES DE DADOS

NATAL/RN
MAIO/2023

SUMÁRIO

APRESENTAÇÃO GERAL	2
Ian Mendes	3
OBJETIVO	3
INTRODUÇÃO	3
METODOLOGIA	3
PLANEJAMENTO	4
ESCOLHAS E RESULTADOS	4
CONCLUSÃO	4
Marcone Rafael	5
OBJETIVO	5
INTRODUÇÃO	5
METODOLOGIA	7
PLANEJAMENTO	8
ESCOLHAS E RESULTADOS	9
Gráfico sobre influencia no índice de felicidade para cada país em 2015:	11
Gráfico sobre influência no índice de felicidade para cada país em 2021:	12
CONCLUSÃO	12
Wellder Bernardo	15
OBJETIVO	15
INTRODUÇÃO	15
METODOLOGIA	16
ANÁLISE E JUSTIFICATIVA	16
PLANEJAMENTO	17
ESCOLHAS E RESULTADOS	19
CONCLUSÃO	22
APÊNDICES	23
Marcone	23
Código do gráfico de 2015:	23
Código do gráfico de 2021:	23
REFERÊNCIAS	24

APRESENTAÇÃO GERAL

Este relatório apresenta três análises distintas, cada uma delas com sua própria importância, método e objetivo. Embora sejam diferentes, todas têm em comum a capacidade de gerar reflexões significativas sobre seus respectivos temas e de responder às perguntas que motivaram sua realização.

A primeira análise aborda a loteria Mega-Sena, investigando seus padrões e resultados. Utilizando um método específico, busca-se compreender os fatores que podem influenciar os números sorteados e oferecer insights relevantes para os jogadores e entusiastas do jogo. A conclusão desta análise fornecerá uma visão mais clara sobre as probabilidades e estratégias envolvidas na Mega-Sena.

A segunda análise concentra-se em como fatores iguais podem ter efeitos diferentes na felicidade de diferentes povos. Por meio de um estudo comparativo, busca-se compreender como variáveis como cultura, economia e sistemas de suporte social influenciam a percepção de felicidade em diferentes países. Essa análise permitirá uma compreensão mais profunda dos elementos que contribuem para a felicidade em diferentes contextos sociais e culturais.

A terceira análise examina como o investimento realizado por um jogador afeta o resultado final de um time de futebol em um campeonato. Com base em dados estatísticos e desempenho esportivo, procura-se estabelecer uma relação entre o investimento financeiro realizado por um jogador e a performance geral da equipe. A conclusão desta análise fornecerá insights valiosos para gestores de clubes esportivos e estrategistas no mundo do futebol.

Em resumo, cada análise proposta neste relatório é única em seu escopo, método e objetivo, mas todas são capazes de gerar reflexões significativas e apresentar conclusões específicas sobre os respectivos temas abordados.

Aluno: Ian Mendes Fernandes do Nascimento

OBJETIVO

O objetivo deste relatório é discorrer sobre o processo de elaboração da representação visual de dados selecionados a partir de um banco de dados referente à modalidade lotérica Mega-Sena. A visualização foi orientada por objetivos de análise. A partir da discussão aqui apresentada, deseja-se propor possibilidades de estudo e visualização do banco aqui mencionado.

INTRODUÇÃO

A Mega-Sena é a maior modalidade lotérica do Brasil. Com prêmios alcançando o valor de mais de 200 milhões de reais, é realizada pela Caixa Econômica Federal desde 1996 em substituição à antiga Sena e consiste na aposta de entre 6 a 20 números em uma tabela de 0 a 60, chamada volante. Na data de elaboração deste relatório, um bilhete de 6 números custa o valor de R\$5,00 (reais), enquanto o de 20 encontra-se a R\$193.800,00 (reais). Os sorteios são realizados duas vezes por semana, às quartas e aos sábados, e os prêmios são distribuídos para os jogadores que acertarem 4 (quadra), 5 (quina) ou 6 valores do volante. A loteria não é considerada a maior do país apenas pelo valor arrecadado, mas também pela diversidade de modalidades de apostas e sorteios, como a Mega da Virada, que soma a arrecadação acumulada de prêmios sem ganhadores e 5% de todos os sorteios do ano, e as Mega Semanas.

Nem todo o dinheiro vai para a premiação, no entanto, a qual soma 43,35% do valor acumulado. O restante é direcionado a diversas instituições, como a Seguridade Social, o Fundo Nacional da Cultura (FNC), o Fundo Penitenciário Nacional (FUNPEN) e o Fundo Nacional de Segurança Pública (FNSP), entre outros.

O banco de dados aqui estudado encontra-se na plataforma Kaggle, onde foi postado há 5 anos no endereço [\[Mega-Sena dataset\]](#) pelo usuário Vinícius Botelho Bizarri. Ele agrega informações comprimidas no intervalo de 11 de Março de 1996 até 25 de Julho de 2018, totalizando 2062 sorteios e seu conteúdo consiste em 22 colunas e 2278 linhas.

METODOLOGIA

A etapa inicial consistiu na pesquisa de uma base de dados que fosse, concomitantemente, alinhada aos interesses pessoais do aluno e estruturada o suficiente de modo a facilitar o tratamento de dados. Outros tópicos foram levantados como possibilidades,

porém, a ausência de informação e a fragmentação do banco de dados por categoria (por exemplo, uma das opções possuía arquivos .csv diferentes para ocorrências de um evento por ano) levaram à escolha do tema: Mega-Sena.

Em posse das informações e após uma breve análise da estrutura do banco de dados, seguiu-se uma reflexão sobre qual seria a questão motivadora a ser respondida por meio da visualização. Depois, foi gerada uma tabela esquematizando possibilidades de representação por variáveis e feito um esboço inicial, propondo a criação de um Mapa de Calor. Esse tipo de gráfico, no entanto, foi descartado por razões expressas na seção “*Escolhas e Resultados*”. Novas opções de gráficos foram estudadas por meio da biblioteca virtual [The Data Visualisation Catalogue](#). Por fim, escolheu-se o tipo de gráfico *Bubble* com histogramas marginais.

A plataforma Google Colab foi utilizada para o tratamento e manipulação de informações e geração de gráficos por meio da biblioteca *Pandas*, em Python. Os gráficos foram programados utilizando a biblioteca *Plotly* e, para um incremento final, os valores referentes ao histograma marginal superior, o qual relaciona a quantidade total de ganhadores por estado, foi utilizado a aplicação web para edição vetorizada *Vectr*.

Todas as etapas, com exceção do incremento final, referentes à programação e criação da visualização encontram-se [neste endereço Colab](#). Os dados encontram-se [neste](#) endereço no drive.

PLANEJAMENTO

Qual a distribuição da quantidade de vencedores da Mega-Sena (acerto de 6 números) e seus respectivos prêmios pelos estados brasileiros desde o início da loteria até o ano de 2018? Com essa pergunta, iniciamos o processo de planejamento. O objetivo é relacionar quatro variáveis: quantidade de vencedores, valor de rateio dos prêmios, estado brasileiro e unidade de tempo. O banco de dados original possui 22 colunas, no entanto, apenas as quatro descritas abaixo foram separadas para análise:

- Data Sorteio: data de realização do sorteio, no formato dia/mês/ano;
- Ganhadores_Sena: quantidade de ganhadores totais em uma determinada data;
- UF: estado dos ganhadores em uma determinada data;
- Rateio_Sena: quanto cada ganhador recebeu individualmente.

Após estudo do banco de dados, percebeu-se algumas características problemáticas: como são relativamente poucos os sorteios com ganhadores, a coluna “*Ganhadores_Sena*”

possuía grande quantidade de valores nulos; em decorrência do fato anterior, a coluna “UF” estava com vários valores vazios; por fim, “Ganhadores_Sena” foi preenchida em função da data, e não do estado do ganhador.

Foi decidido que seria utilizada a biblioteca *Pandas* com a linguagem *Python* para a manipulação da base de dados e realização de sua limpeza. A fim de resolver o primeiro problema citado no parágrafo acima, todas as linhas que apresentavam o valor 0 em “Ganhadores_Sena” foram descartadas, consequentemente, descartando as linhas com valor NaN em “UF” e solucionando o segundo problema. Já a quantidade de ganhadores em função do estado e da data foi encontrada por meio da criação de uma nova coluna, “Ocorrências”, por meio do agrupamento e contagem da repetição dos UF em cada data específica.

As especificidades de programação encontram-se no [link Colab](#), abaixo de *Tratamento de Dados*. Desse modo, alguns valores foram alterados: a variável “ganhadores” transforma-se na coluna “ocorrências” e passa a ter o intervalo 1 a 6 ganhadores por estado ao invés de 0 a 17 ganhadores totais por data.

Variável	Tipo de dado	Valores	Representação 1		Representação 2	
			símbolo	representação visual	símbolo	representação visual
Data de Sorteio	quantitativo contínuo	de 11/03/1996 a 25/07/2018	ponto	posição horizontal em uma linha	área	posição vertical
Ganhadores	quantitativo discreto	1-6	área	área do círculo	área	escala de cor
UF	qualitativa nominal	Todos os 27 estados brasileiros	ponto	posição vertical em uma linha	área	posição horizontal
Rateio (Valor)	quantitativo contínuo	0 - 205.329.753,89	área	escala de cor	área	área de um círculo

Seguiu-se a criação de rabiscos de potenciais gráficos após o estudo de tipos de gráficos em sala de aula. Por interesse próprio, foi estudada a possibilidade da realização de um Mapa de Calor. O rascunho (realizado antes da decisão de deletar todos os zeros) encontra-se abaixo:

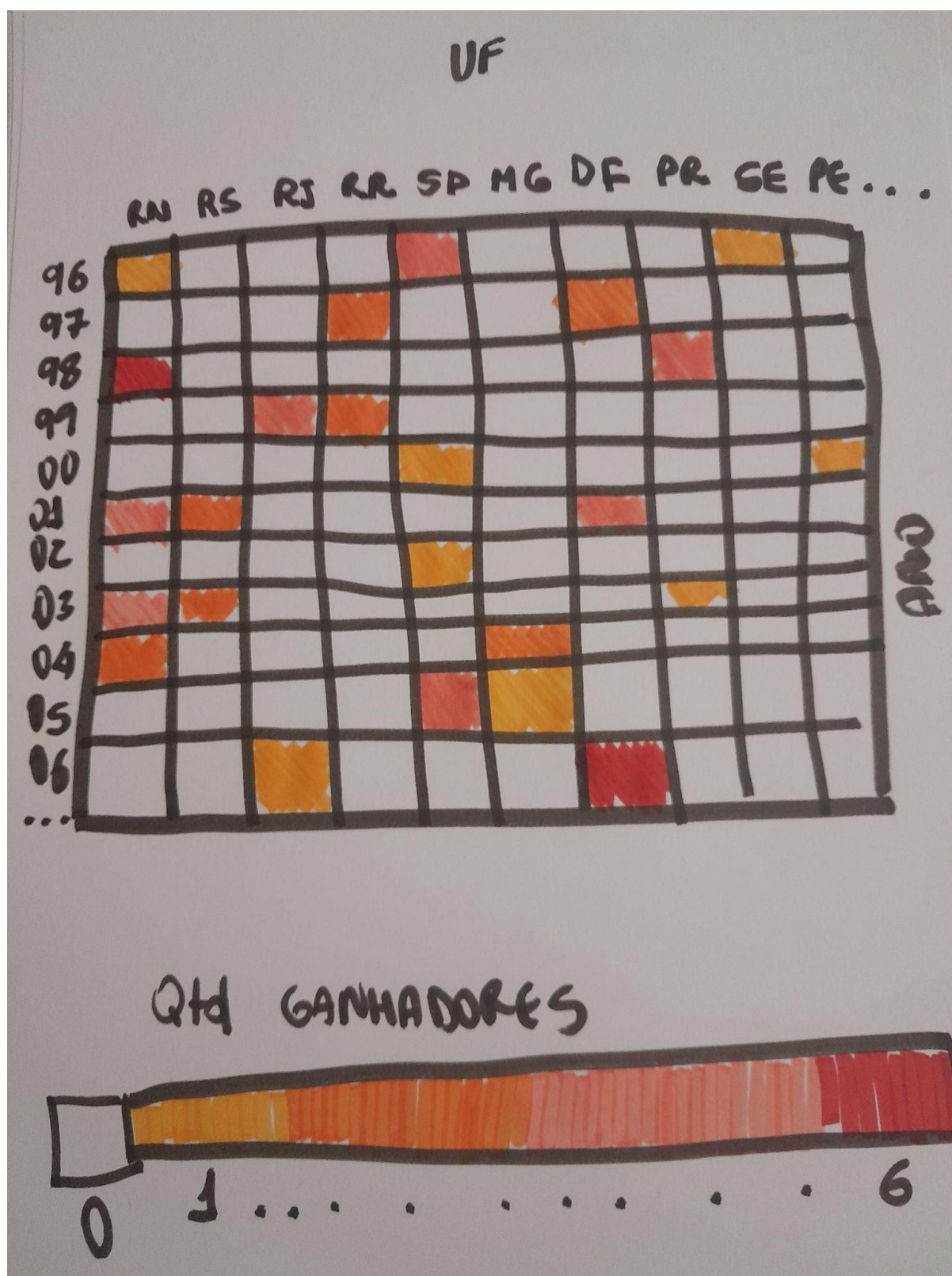


Imagem 1 - Rabisco de esquematização inicial. Autoria própria

Pelos rabiscos iniciais, ficou evidente a dificuldade de inserir uma quarta variável (neste caso, o valor de rateio) de forma coerente. A data torna-se um outro desafio, pois se fosse inserida em dias do ano, geraria uma quantidade considerável de linhas no mapa,

complexificando-o o suficiente para torná-lo quase incompreensível. Caso a variável fosse agregada em anos, por exemplo, perderia-se qualidade de informação, pois ela seria menos acurada em sua distribuição. A retirada dos valores nulos em ganhadores diminuiria a quantidade de linhas nas datas, mas não o suficiente: de 2278 teríamos 678 linhas.

Foi decidido que, por acurácia, o gráfico seria programado utilizando a biblioteca *Plotly*, por flexibilidade e familiaridade. Após descartar o tipo Mapa de Calor como opção, percebeu-se que um gráfico tipo *Bubble* poderia representar, de forma mais compreensível, as quatro variáveis com menos perda de informação.

ESCOLHAS E RESULTADOS

Gráficos do tipo *Bubble* possuem uma boa flexibilidade para passar informações com os mais diversos tipos de variáveis. A área da representação visual de escolha, assim como a cor e a posição em que se encontra convergem, juntos, a relação entre quatro variáveis. Tendo isso em mente, foram exploradas as possibilidades, programando combinações pelo Colab.

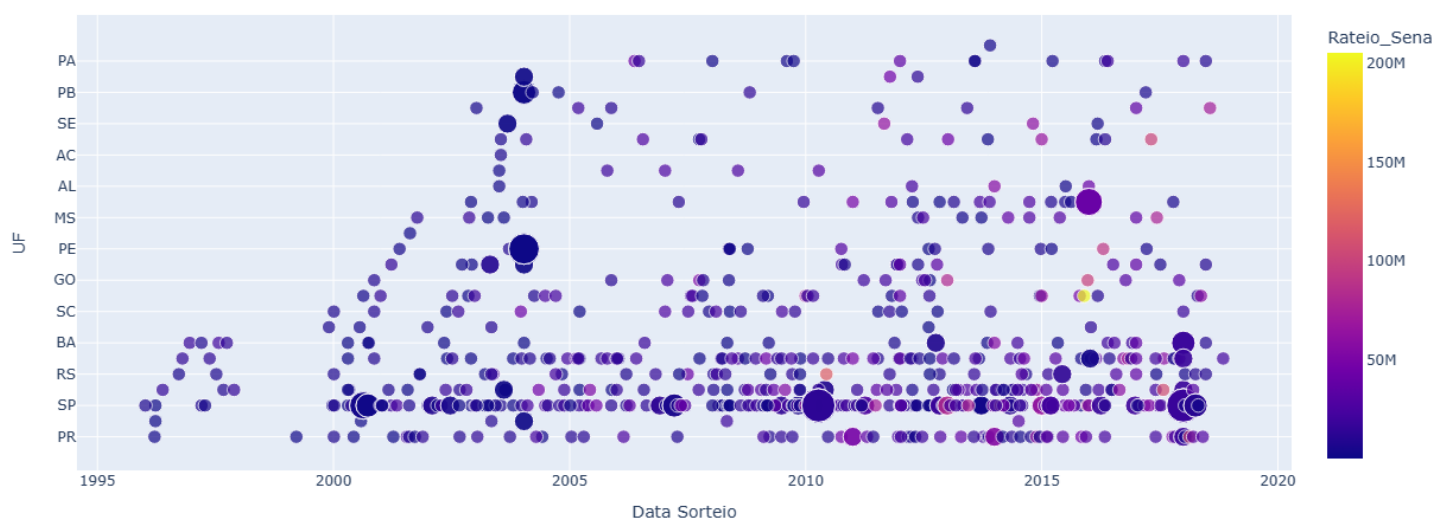


Imagem 2 - Possibilidade de gráfico programado. Autoria própria.

Este modelo deu suporte à proposta da utilização de um gráfico tipo *Bubble*, pois comunica visualmente uma ideia geral da data e estado dos maiores prêmios. Como é um gráfico interativo, ao passar o cursor em cada círculo somos capazes de ver os dados completos daquele ponto. No entanto, visualmente, o desafio da perda de informação persistia. Com exceção do valor de prêmio, não havia padronização visível da quantidade de ganhadores por estado e pelo tempo, relações importantes para a análise.

Por meio do estudo da biblioteca utilizada, descobriu-se a possibilidade da criação de gráficos marginais. Por meio de testes, chegou-se à conclusão que histogramas no eixo x - relacionando quantidade de ganhadores à UF - e no eixo y - relacionando quantidade de ganhadores à data - convergiam as questões trazidas pela pergunta motivadora de forma visual, sem depender da interatividade on-line.

Com retoques finais feitos na aplicação web *Vector* e uma pequena alteração na escala de cores para favorecer o contraste, com alguma perda de qualidade (o arquivo original de maior qualidade encontra-se em formato .pdf), o resultado final foi alcançado e encontra-se abaixo.

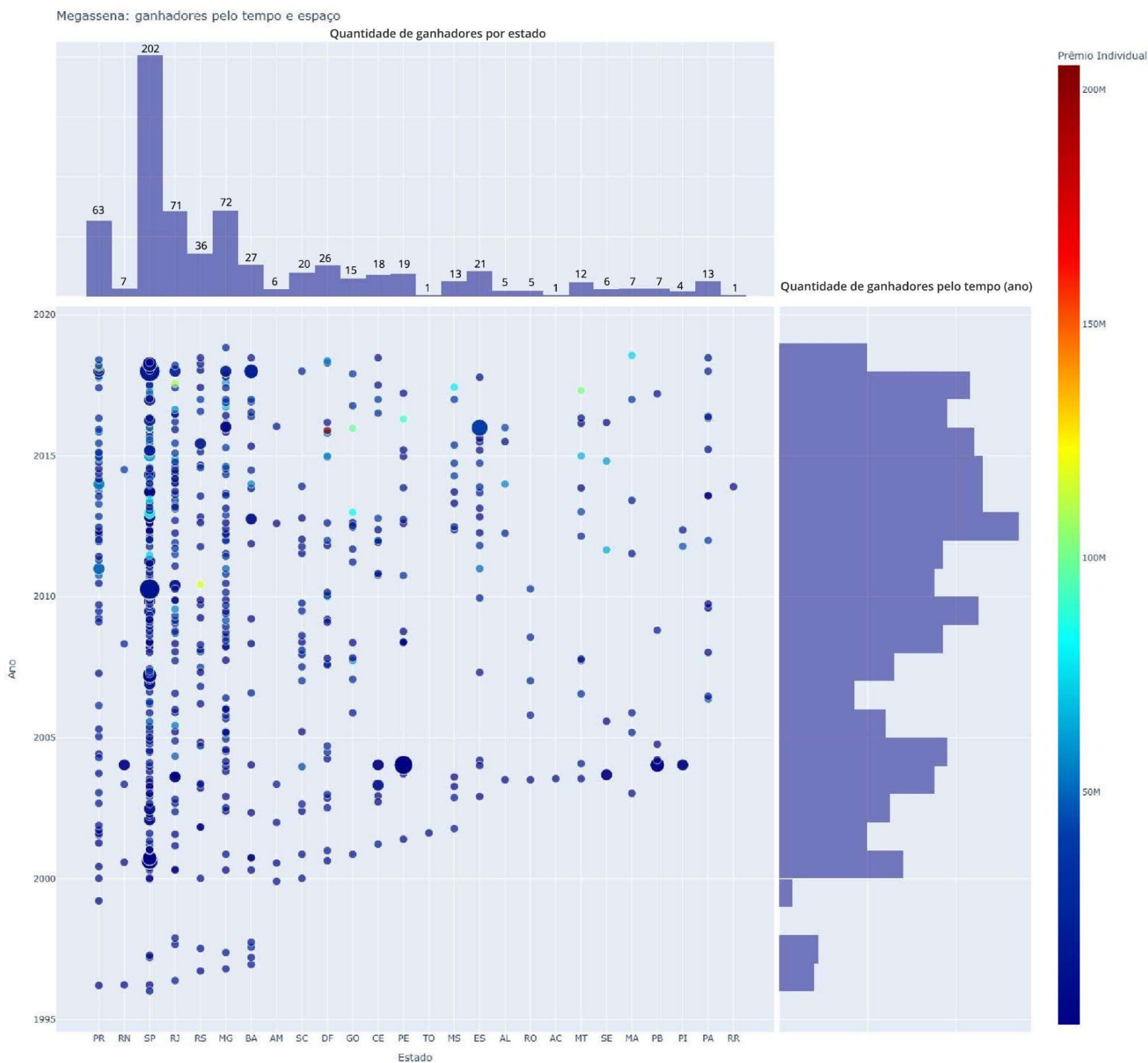


Imagem 3- Gráfico final : Ian. Autoria própria.

CONCLUSÃO

O estado de São Paulo concentra a maior quantidade total de ganhadores, totalizando 202, enquanto Roraima, Tocantins e Mato Grosso disputam a última colocação com 1 ganhador em toda a história da loteria. Enquanto em 1998 ninguém ganhou, 54 apostadores

foram premiados em 2012, sendo esse o ano com a maior quantidade de prêmios. O maior prêmio de todos foi dado a um morador do Distrito Federal em 2015, R\$ 205.329.753,89. A soma dos prêmios totais do início da Mega-Sena até os anos 2000 é menor do que após a virada do milênio.

O que esses dados apontam? Eles dirigem nossa atenção à possíveis correlações com outros fatores, geográficos, talvez históricos e econômicos. Será que há relação entre a quantidade de ganhadores e a população de um estado? Se sim, como isso se expressa estatisticamente? O aumento do valor do prêmio acumulado indica maior crescimento da popularidade da loteria a partir dos anos 2000. A que isso se deve? Será que o poder aquisitivo dos residentes de cada estado se expressa na quantidade de ganhadores?

Uma boa esquematização visual de dados auxilia na criação de perguntas a fim de se aprofundar no tema proposto. A partir deste trabalho, concluo que a criação de representações visuais são pequenos degraus na compreensão de fenômenos e suas relações.

OBJETIVO

O objetivo deste relatório é realizar uma análise abrangente sobre como os mesmos fatores afetam o índice de felicidade do World Happiness Report da ONU de maneira diferente em cada país. Através de pesquisas realizadas em 2015 e 2023, examinaremos a influência da expectativa de vida, PIB per capita, percepções de corrupção e a liberdade para fazer escolhas de vida nos países. O objetivo é compreender as possíveis relações e interações entre esses fatores socioeconômicos e a longevidade da população, contribuindo para um maior entendimento dos determinantes que influenciam o indicador de felicidade.

INTRODUÇÃO

No cenário contemporâneo, o índice de felicidade desempenha um papel fundamental como indicador do desenvolvimento e bem-estar das sociedades. Problemas como a depressão têm um impacto significativo na saúde mental das pessoas. Portanto, compreender os fatores que afetam a felicidade da população é essencial para a formulação de políticas públicas eficazes e estratégias de desenvolvimento sustentável. Nesse contexto, esta análise propõe uma investigação detalhada, levando em consideração a expectativa de vida, o PIB per capita, as percepções de corrupção e a liberdade para fazer escolhas como elementos que podem influenciar essa expectativa.

A expectativa de vida é um indicador importante que reflete a saúde da população de um país e pode ter um impacto significativo na felicidade das pessoas. Quanto maior a expectativa de vida, geralmente é um indicativo de uma população saudável e pode contribuir positivamente para a felicidade.

O PIB per capita, como medida de riqueza econômica de um país, está frequentemente relacionado a melhores condições de vida e acesso a serviços de saúde. Isso pode ter um efeito positivo na felicidade das pessoas, uma vez que um maior PIB per capita está associado a melhores oportunidades e qualidade de vida.

Além disso, as percepções de corrupção podem afetar a alocação de recursos públicos, comprometendo os investimentos em áreas cruciais como saúde e educação. Isso pode ter um impacto negativo na felicidade da população, uma vez que a corrupção pode prejudicar a confiança nas instituições e na eficácia das políticas governamentais.

Por outro lado, o nível de liberdade de uma nação pode refletir a capacidade dos indivíduos de exercerem seus direitos e liberdades fundamentais, influenciando diretamente seu bem-estar subjetivo e, consequentemente, o índice de felicidade. Um maior grau de liberdade pode estar relacionado à sensação de autonomia, escolha e participação na tomada de decisões pessoais e coletivas, aspectos essenciais para a satisfação e realização pessoal. Além disso, um ambiente de maior liberdade pode promover a diversidade, a criatividade, a expressão individual e a igualdade de oportunidades, fatores que têm sido associados a níveis mais elevados de felicidade em estudos anteriores. Portanto, é importante investigar a relação entre o nível de liberdade e o índice de felicidade, a fim de compreender melhor como esse fator contribui para o bem-estar das pessoas em diferentes países.

Em suma, a expectativa de vida, o PIB per capita, as percepções de corrupção e a liberdade são fatores importantes que podem influenciar a felicidade das pessoas em um país, e compreender essas relações é essencial para promover um maior bem-estar social.

Ao investigar essas variáveis e suas possíveis inter-relações, meu objetivo é obter uma compreensão mais profunda dos fatores que moldam a felicidade das pessoas em diferentes países. Por meio de uma análise crítica e fundamentada dos dados disponíveis, buscarei identificar associações e tendências significativas. Essas informações serão valiosas para embasar decisões e ações direcionadas à promoção da saúde e qualidade de vida.

Ao entender melhor os determinantes da felicidade, poderemos desenvolver políticas públicas mais eficazes e estratégias de desenvolvimento sustentáveis, que

abordem os fatores-chave que influenciam a satisfação e bem-estar das pessoas. Isso contribuirá para a criação de sociedades mais saudáveis, resilientes e felizes.

A análise crítica dos dados disponíveis é essencial para obter insights significativos e confiáveis. Portanto, meu trabalho será baseado em uma abordagem cuidadosa e fundamentada, utilizando métodos de pesquisa adequados e levando em consideração a validade e confiabilidade dos dados analisados.

METODOLOGIA

Para realizar esta análise, adotei uma abordagem metodológica baseada na coleta de dados e na utilização de técnicas de visualização. Inicialmente, busquei informações relevantes a partir do World Happiness Report, referentes aos anos de 2015 e 2023. Esses relatórios foram obtidos através do site [World Happiness Report \(kaggle\)](https://www.kaggle.com/datasets/stevenbarlow/world-happiness-report), e fornecem dados abrangentes sobre a felicidade e bem-estar em diversos países ao redor do mundo, inclusive os que eu utilizei que revelam para cada país o quanto cada variável faz a diferença no índice final de felicidade.

Utilizei a plataforma Google Colab e importei a biblioteca Pandas para a manipulação e análise dos dados, assim como as bibliotecas Matplotlib.pyplot e Matplotlib.cm para a criação dos gráficos do tipo scatter plot. Esse tipo de gráfico é ideal para visualizar a relação entre variáveis quantitativas, o que me permitiu explorar as relações entre a expectativa de vida, o PIB per capita, as percepções de corrupção, liberdade e o índice de felicidade.

Após a importação dos dados do World Happiness Report para o ambiente do Google Colab, realizei uma limpeza e organização dos dados, tratando eventuais valores ausentes ou inconsistentes. Em seguida, selecionei as variáveis relevantes para a análise, como a expectativa de vida, o PIB per capita, as percepções de corrupção e a liberdade.

Utilizando essas variáveis, criei dois gráficos scatter plot distintos, um para o ano de 2015 e outro para o ano de 2023. Esses gráficos me permitiram visualizar a dispersão dos dados e identificar possíveis tendências ou correlações entre as variáveis analisadas.

Por meio dessa metodologia, busquei explorar visualmente os dados e fornecer insights sobre a relação entre a expectativa de vida e os fatores socioeconômicos considerados neste estudo. Essa abordagem me permitiu observar possíveis padrões ou comportamentos nos dados, contribuindo para uma compreensão mais aprofundada dos determinantes da longevidade nos países analisados.

PLANEJAMENTO

Partindo da pergunta "Dentre a variação da expectativa de vida dos países, o PIB per capita, as percepções de corrupção e a liberdade, quanto a contribuição de cada uma dessas variáveis afeta de forma diferente o índice de felicidade em cada país?", o plano de análise incluirá o uso de scatter plots como ferramenta de visualização dos dados. O objetivo é identificar possíveis relações e tendências entre essas variáveis socioeconômicas e a expectativa de vida com o índice de felicidade de cada país nos anos de 2015 e 2023. A análise crítica dos scatter plots permitirá observar visualmente as associações entre as variáveis e fornecer insights significativos sobre os fatores que influenciam a felicidade. Ao examinar as relações entre essas variáveis, será possível obter uma compreensão mais aprofundada dos determinantes da felicidade em diferentes contextos nacionais.

O planejamento propunha a criação de dois gráficos scatter plot, um para cada ano, em que cada ponto representa o índice de felicidade de um país. As coordenadas dos pontos correspondem à expectativa de vida e ao PIB per capita, respectivamente. Além disso, seria utilizado um esquema de cores diferentes para representar os níveis de liberdade, atribuindo tonalidades específicas com base nas escalas dessas variáveis. Adicionalmente, o tamanho dos pontos poderia variar de acordo com as percepções de corrupção. Como ilustra a representação 01 na tabela a seguir.

Variável	Tipo de dado	Possíveis valores	Representação 01	Representação 02

			<i>Símbolo 1</i>	<i>Características visuais 1</i>	<i>Símbolo 2</i>	<i>Características visuais 2</i>
índice de felicidade de cada país	Categoria	138 países	ponto	posição	ponto	rótulo
PIB per capita	quantitativo	De 0 a 2 <i>(usam essa escala para evidenciar as diferenças)</i>	ponto	rótulo	linha	posição
Liberdade	quantitativo	De 0 a 1	Área	Faixa de cor	ponto	cor
expectativa de vida saudável	quantitativo	De 0 a 1	ponto	rótulo	rótulo	Faixa de cor
Percepções de corrupção	quantitativo	De 0 a 1	Área	Tamanho	Área	Tamanho

ESCOLHAS E RESULTADOS

Primeiramente devido às diferenças na abordagem dos dados de 2023 e 2022 com relação às escalas usadas em 2015, devido a isso fiz o segundo gráfico sobre o ano de 2021.

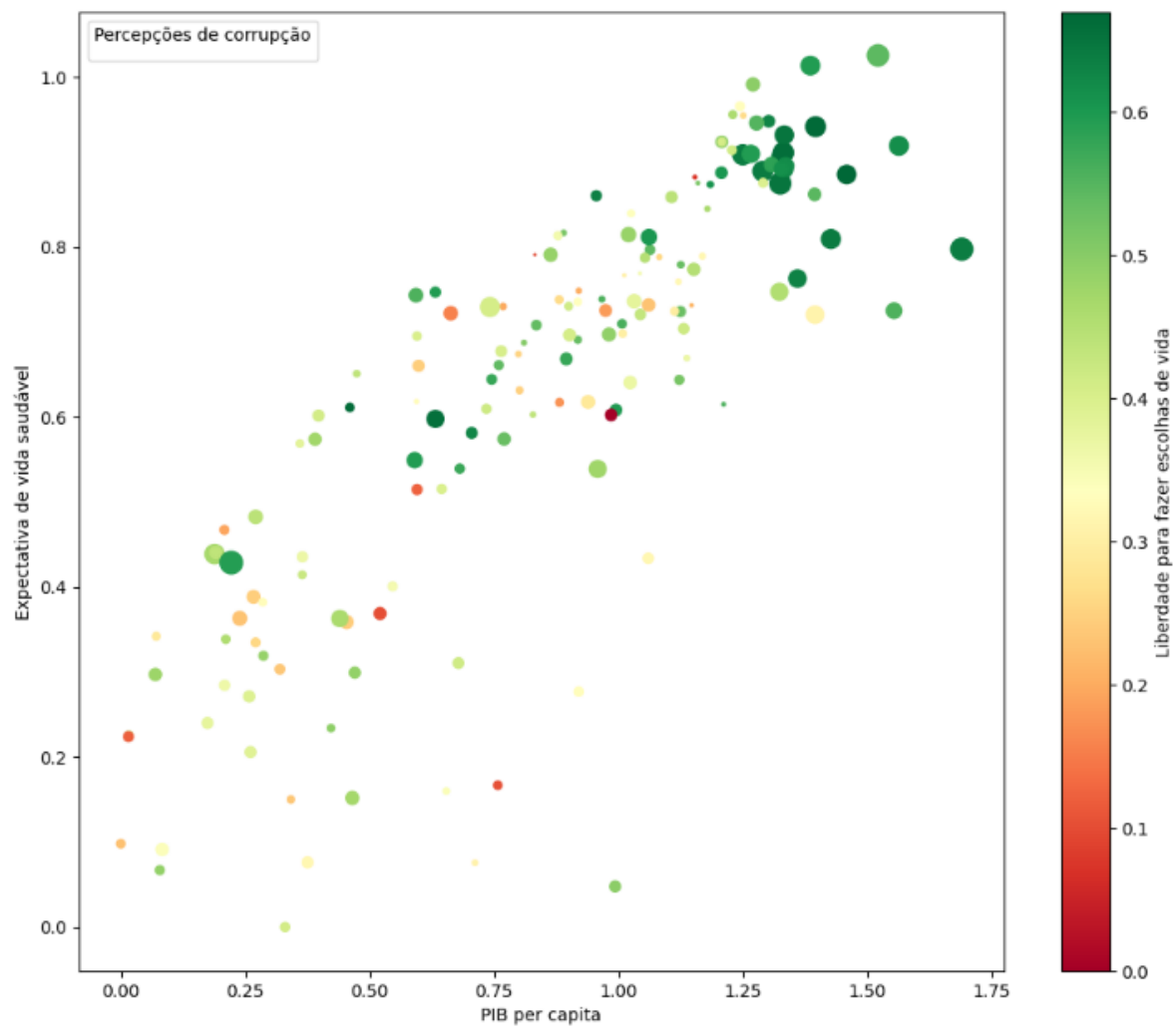
Para executar esse plano, foram utilizadas as bibliotecas Pandas, Matplotlib.pyplot e Matplotlib.cm, disponíveis no ambiente do Google Colab. Os dados necessários foram obtidos a partir do World Happiness Report dos anos de 2015 e 2023, acessados por meio do site [kaggle.com/World Happiness Report](https://kaggle.com/WorldHappinessReport).

Após a preparação dos dados, foram criados os dois gráficos scatter plot, um para cada ano, seguindo as especificações estabelecidas no esboço do planejamento. Esses gráficos permitiram visualizar a relação entre as variáveis e identificar possíveis padrões ou correlações.

O resultado dessa análise forneceu insights sobre a variação do índice de felicidade em relação a expectativa de vida e, PIB per capita, percepções de corrupção e nível de liberdade nos países analisados. Essas informações contribuíram para um maior entendimento dos determinantes socioeconômicos que impactam na felicidade da população e podem subsidiar a formulação de políticas públicas e estratégias de desenvolvimento mais eficazes.

Os gráficos a seguir foram construídos com base no planejamento que foi anteriormente explicado mostrando o quanto cada variável influencia na felicidade de cada país.

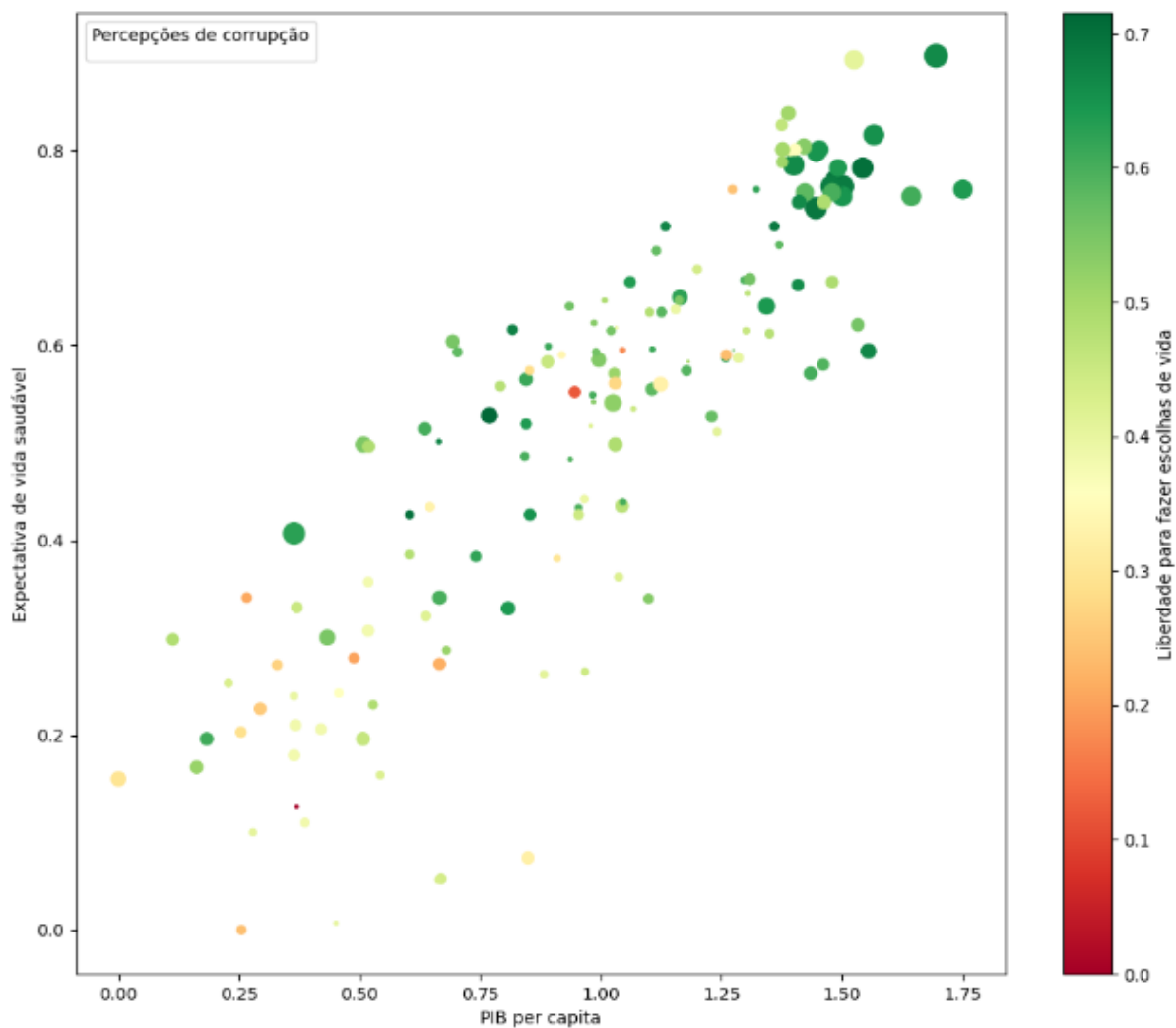
Gráfico sobre *influencia* no índice de felicidade para cada país em 2015:



Legenda:

- Cada ponto representa um país
- O tamanho dos pontos representam a influência negativa da corrupção nos países
- OBS: Os dados são a influência de cada variável na felicidade dos países

Gráfico sobre *influência* no índice de felicidade para cada país em 2021:



Legenda:

- Cada ponto representa um país
- O tamanho dos pontos representam a influência negativa da corrupção nos países
- OBS: Os dados são a influência de cada variável na felicidade dos países

Ambos os códigos estão disponíveis em [colab](#) lembrando que é necessário fazer upload dos [dados](#) no colab para que o código funcione corretamente.

CONCLUSÃO

Quando nos deparamos com os gráficos que mostram a influência das variáveis na felicidade, uma constatação importante se revela: cada país possui suas particularidades culturais que exercem um poderoso impacto nesse sentido. É como se

a cultura de cada povo moldasse e tempera-se a busca pela felicidade de forma única, de maneira que o que traz contentamento para a maioria de uma cultura pode não surtir o mesmo efeito em outros povos, levando em conta também uma série de outros fatores.

É inegável que a cultura exerce uma influência marcante nas expectativas, valores e prioridades das pessoas. Ela nos guia, nos oferece parâmetros e referenciais que nos ajudam a compreender o que é a felicidade em nossa própria vivência. Porém, é preciso ter a sensibilidade de perceber que essa experiência culturalmente construída não é universal, e que cada sociedade traz consigo suas próprias visões de mundo e aspirações.

Ao refletirmos sobre a relação entre as variáveis socioeconômicas e a felicidade, é fundamental considerar a influência desses diferentes contextos culturais. Afinal, as crenças, normas e tradições variam significativamente de uma cultura para outra, e isso impacta diretamente o que nos faz felizes. O que é valorizado em uma cultura pode não ter o mesmo apreço em outra.

Contudo, é importante destacar que a cultura não é o único fator determinante nesse processo. Há uma teia complexa de influências que se entrelaçam: o contexto histórico, as condições econômicas, as políticas governamentais, o acesso a recursos e tantos outros aspectos. Esses fatores interagem com as características culturais e contribuem para moldar as expectativas e as fontes de felicidade em cada país.

Nessa perspectiva, compreender a relação entre cultura, variáveis socioeconômicas e felicidade requer um olhar sensível e uma abordagem que reconheça a pluralidade de experiências. Devemos considerar que aquilo que nos traz contentamento em nossa própria cultura não é uma verdade universal, e que as soluções para promover a felicidade em cada sociedade devem levar em conta essas nuances culturais.

Portanto, ao analisar os gráficos e os dados que revelam as diferenças de impacto das variáveis na felicidade, devemos ter em mente a importância de uma abordagem humanizada e respeitosa das diversas culturas. É através desse

entendimento que poderemos construir políticas públicas e estratégias de desenvolvimento que estejam verdadeiramente alinhadas com a busca pelo bem-estar e a felicidade dos indivíduos em suas mais variadas vivências culturais.

OBJETIVO

Partindo como tema da análise dados do Campeonato Brasileiro de futebol, seguindo um recorte de 10 anos entre 2009 e 2018, com o objetivo de comparar o desempenho esportivo, dos clubes que se classificaram para as competições internacionais e dos clubes que foram rebaixados da Série A para a Série B, e com isso identificar quais clubes nesses dois grupos tiveram sucesso ou insucesso comparado ao seu nível de investimento por jogador e sua colocação final.

INTRODUÇÃO

Organizado pela CBF, Confederação Brasileira de Futebol, o Campeonato Brasileiro é a principal competição futebolística do Brasil. Por ela, são indicados os representantes brasileiros nas competições internacionais, seja a Copa Libertadores ou a Copa Sul-americana.

A competição durante a história possui diversos nomes desde o ano de 1959 com a Taça Brasil, em 1967 passou a chamar-se de Torneio Roberto Gomes Pedrosa, em 1971 tornou-se o Campeonato Nacional de Clubes e em 1989 finalmente consolidou o nome Campeonato Brasileiro de Futebol. Assim como seus nomes, a liga também possuiu vários formatos, sendo eles às vezes tocados de ano em ano, ocorrendo uma característica falta de padronização em seu sistema de disputa, com isso em diversas temporadas não ocorriam acessos ou decessos entre divisões. Dentre esses formatos ocorreram sistemas eliminatórios (1959-1968), sistemas mistos de grupos (1968-2002) e o sistemas de pontos corridos que é o adotado até hoje com todos os clubes enfrentando-se em turno e retorno e com critérios de desempate baseado em números de vitórias, saldo de gols, gols marcados e até cartões amarelos.

Recentemente o até o ano de 2016 os quatro melhores colocados se classificavam para a libertadores, porém, com a saída dos clubes Mexicanos da mesma a quantidade de vagas foram aumentadas dentre os países participantes, com isso, Brasil passou de quatro para seis vagas, logo, essa mudança será identificada na análise.

METODOLOGIA

Para desenvolver a análise foram utilizados dados compartilhados na plataforma Kaggle pelo usuário André Vinícius onde alguns desses dados tratados por ele foram retirados do website Transfermarkt, que é uma plataforma com foco em da informações sobre transações de jogadores a respeito de valores.

O dataset utilizado é composto de 13 variáveis sendo elas, anos, posição final do clube no campeonato além de informar o próprio clube que ocupou aquela colocação, além da quantidade de vitórias, derrotas, empates, gols feitos, gols sofridos, saldo de gols que é a diferença entre as quantidades anteriores, quantidade de jogadores, idade média do elenco, quantidade de estrangeiros, valor total do time e por fim a média de valor entre a quantidade de jogadores e o valor total da equipe. Posteriormente as variáveis de interesse para a análise foram separadas e tratadas para formular a tabela de dados, com isso, surgiu a necessidade de criar a variável pontos, que nada mais é do que a soma do desempenho de vitórias e empates das equipes, onde a vitória vale 3 e o empate vale 1, ocorrendo assim uma manipulação de dados extra no desenvolvimento.

Para criação e desenvolvimento do gráfico foi utilizado a plataforma de desenho vetorial Vectr que é um software de multiplataformas além da versão web que foi a escolhida para a construção. A escolha se deu pela facilidade e praticidade de manipulação da interface para a criação de gráficos vetoriais e ainda por ser gratuito.

ANÁLISE E JUSTIFICATIVA

Seguindo com a pergunta para a análise “Qual o desempenho esportivos dos clubes nas 5 primeiras posições e nas últimas 5 posições relacionada ou valor do elenco por jogador no período de 2009 a 2018 ?”.

Desempenho esportivo nada mais é do que o resultado do time na temporada, neste caso, estaremos medindo apenas o Campeonato Brasileiro pois é o que os

nostros dados nos fornece. Sendo assim, a comparação se dá pela quantidade de pontos e consequentemente a posição final, vale salientar também a comparação da média do valor de elenco por jogador, pois em tese, o preço do jogador é concebido inerente a sua qualidade, de modo que elencos mais caros tem a pressão por resultados melhores, porém, nem sempre isso ocorre por vários fatores, sejam eles, foco em outras competições, treinadores mal sucedidos, lesões, instabilidade relacional entre o grupo, simplesmente sorte, entre outros.

Dito isso, a análise foca em dois grupos distintos da tabela final do Campeonato, os cinco primeiros onde quatro conseguem a classificação para a Libertadores com destaque para o campeão e a ideia de comprar o quinto é para identificar o time que quase conseguiu a classificação, vale ressaltar que a partir de 2017 o quinto colocado passou a clasificarse também, logo, não poderá ocorrer essa identificação. Os 5 últimos têm a mesma ideia, porém em relação ao rebaixamento, onde os quatro últimos são rebaixados e identificando o décimo sexto que se salvou da queda.

PLANEJAMENTO

Dita a análise e suas justificativas partimos para o planejamento do desenvolvimento visando tentar colocar todas as ideias em um gráfico bem estruturado e solidificado. Primeiramente necessita pelo menos de duas ideias iniciais de símbolos e características visuais para cada variável da análise, que são elas, posição final, pontos, anos e média do valor do elenco por jogador. Segue a tabela com as colocações:

Variável	Tipo de Dado	Possíveis Valores	Representação 1		Representação 2	
			Símbolo 1	Característica Visual 1	Símbolo 2	Característica Visual 2
Posição Final	Categoria	1º - 5º 16º - 20º	linha	cores	linha	cores
Pontos	Quantidade	20 - 81	pontos	posição	área	forma radial
Anos	Ordem	2009 - 2018	pontos	posição	área	forma radial
Média de valor por jogador	Quantidade	0,10M - 3,00M (Milhões)	área	tamanho de círculos	ponto	posição

Após relacionar as variáveis separadamente é necessário destacar os dados que serão utilizados na análise por meio de uma tabela de dados. Nela estão organizados em colunas com os anos e essa própria é dividida em duas, uma com os pontos e outra com a média de valor por jogador, e nas linhas estão as posições de interesse. Segue a tabela:

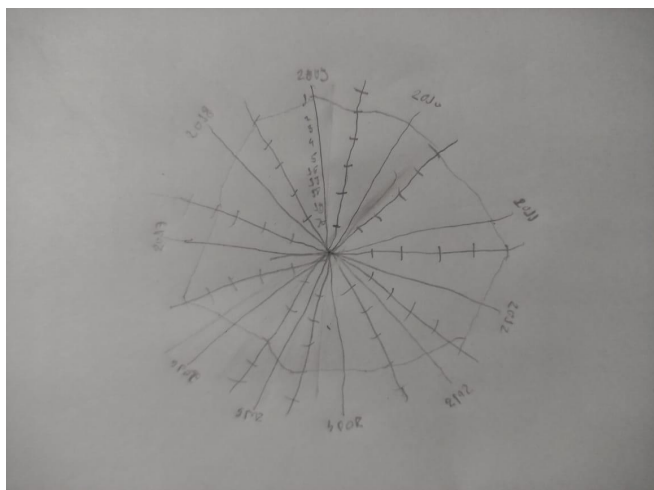
	2009		2010		2011		2012		2013		2014		2015		2016		2017		2018	
1	67	1,44M	71	0,86M	71	1,37M	77	1,11M	76	1,10M	80	1,31M	81	1,47M	80	1,84M	72	1,21M	80	1,84M
2	65	0,94M	69	1,24M	69	0,96M	72	1,40M	65	0,98M	70	2,52M	69	1,34M	71	1,06M	63	1,57M	72	1,40M
3	65	1,56M	68	1,31M	63	1,19M	71	1,32M	64	2,27M	69	0,89M	68	0,78M	71	1,74M	63	1,18M	69	1,21M
4	62	1,53M	63	1,09M	61	1,49M	66	2,18M	61	0,43M	69	2,03M	62	2,11M	62	1,60M	62	0,56M	66	1,15M
5	62	0,89M	60	0,54M	60	1,39M	58	0,80M	59	0,34M	62	1,46M	60	1,18M	59	0,53M	57	1,67M	63	1,42M
16	46	1,36M	42	0,21M	43	1,51M	45	0,22M	45	1,26M	40	0,84M	43	0,25M	45	0,78M	43	0,81M	43	0,75M
17	45	0,54M	42	0,39M	41	0,55M	41	0,35M	44	0,25M	38	0,31M	42	0,23M	43	1,32M	43	0,65M	42	0,66M
18	41	0,14M	37	0,24M	39	0,50M	34	0,91M	44	0,55M	37	0,38M	41	0,43M	37	0,66M	43	0,22M	40	0,53M

19	38	0,46M	33	0,60M	37	0,13M	30	0,31M	37	0,33M	34	0,75M	38	1,30M	31	0,36M	39	0,65M	37	0,72M
20	31	0,52M	28	0,24M	31	0,44M	30	0,36M	20	0,17M	32	0,33M	31	0,12M	28	0,47M	36	0,46M	23	0,34M

Feito isso, já temos os requisitos necessários para a criação de rascunhos e quem sabe mudanças de ideias e alterações proveniente da inserção de características conceituais dos tipos de dados e representações visuais estudados durante as unidades.

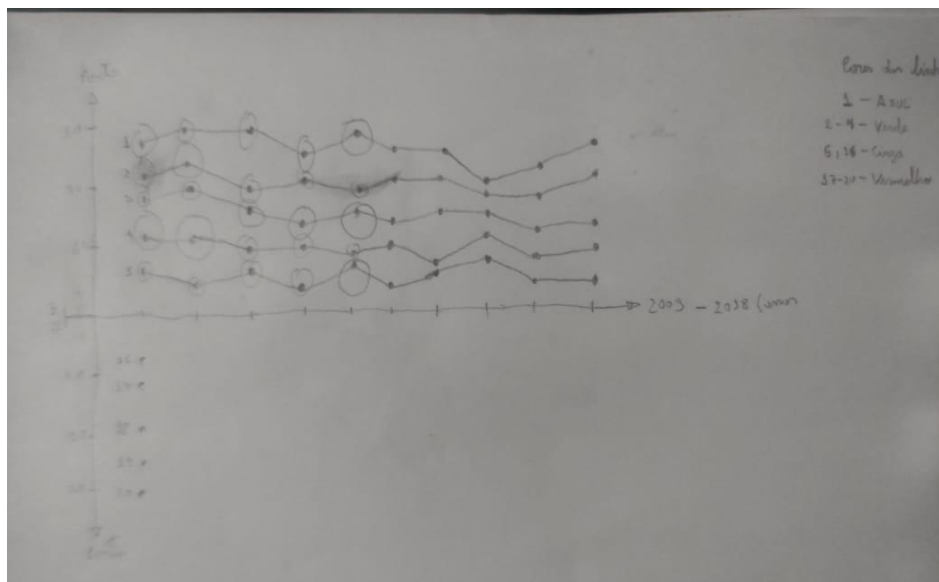
ESCOLHAS E RESULTADOS

Após colocar todas as ideias à mesa e com toda a base de dados de interesse, da-se início a fase de rascunho e produção do gráfico final. Seguindo a representação 2 sugerida na tabela de representações temos um gráfico radial com linhas, a seguir temos um esboço inicial:



Basicamente, o ângulo foi dividido em 10 partes para cada ano e ao lado direito indica a quantidade de pontos, na própria linha dos anos, foi dividida em posições, e as linhas que cortam os arcos representam cada posição em relação aos pontos, ligando uns aos outros. Entretanto, além da falta de espaço para a percepção, não foi possível encaixar a última variável de interesse, desse modo, essa forma de representação foi descartada.

Partindo para a representação 1, temos um gráfico de linha com eixo horizontal deslocado para o centro do do eixo vertical com formas circulares nas posições de interesse. Abaixo temos o esboço do gráfico:



Como já foi citado é um gráfico de linhas com o eixo horizontal deslocado, no vertical temos a representação dos pontos, no horizontal temos os anos, as linhas temos as posições finais e as áreas representam o valor do elenco por jogador, quanto maior o círculo, maior é o valor representado.

Seguindo o embasamento teórico, a respeito da primeira variável, que no caso é a variável de categoria relacionada a posição final, temos uma visualização por cores que nos casos de primeiro lugar é uma cor isolada, no caso de segundo a quarto é uma cor com tonalidades diferente, a quinta e a décima sexta posição uma cor neutra e as últimas quatro posições temos tons de outra cor. Essa maneira de distribuição de cores mapeia muito bem a diferenciação dos objetivos desejados, referenciando entre os eixos e passando a ideia de posição ordenada podendo se relacionar com as outras variáveis de desejo.

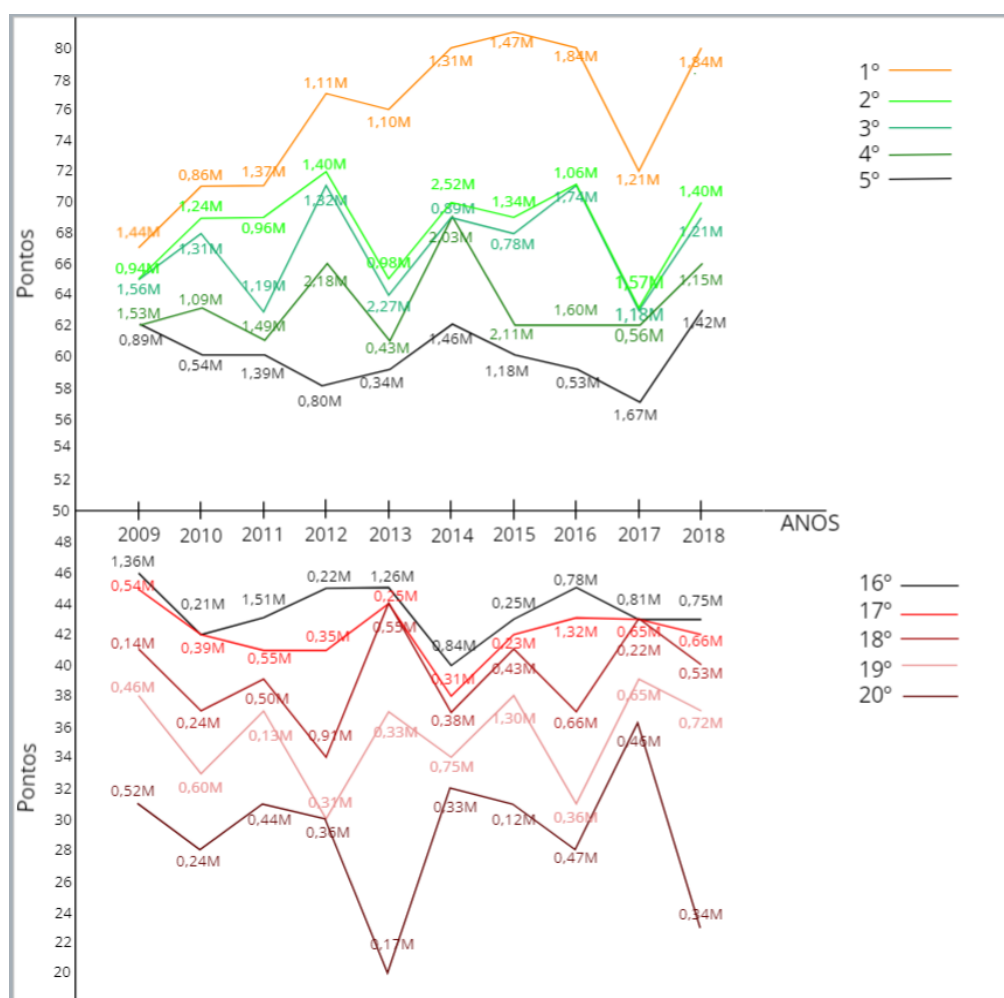
As variáveis de pontos e anos foram as escolhidas para ficar nos eixos vertical e horizontal respectivamente, com um adendo que já foi citado que é o deslocamento do eixo x para que a separação dos grupos distintos da análise ficassem mais incisiva, desse modo a percepção do pelotão de cima tivesse um foco na parte de cima e o pelotão de baixo tivesse esse foco embaixo. Essa manipulação só foi possível pelo

fato da característica dos dados de n passarem uma certa margem, com isso nenhuma linha cortaria o eixo horizontal tornando a escolha bem inteligente passando a sensação de blocos desejada pela análise.

Por fim, a variável média de valor por jogador. Essa variável gera essa informação de cada clube durante aquele ano específico e sendo assim uma das mais importantes da análise por ter o maior caráter comparativo. Porém ocorreu um problema na hora do seu encaixe, a sobreposição de formas tornou-se um problema para o gráfico fazendo com que ficasse muito difícil de perceber alguns pontos das outras variáveis, por esse motivo surgiu a necessidade de tirar as formas circulares e substituir por rótulos informando os valores, desse modo seria possível perceber todas as variáveis.

Segue agora o gráfico final:

GRÁFICO DE DESEMPENHO POR VALOR DE MÉDIO DO BRASILEIRÃO 2009-2018



CONCLUSÃO

Finalizada a apresentação e discussão, é proveniente perceber que a ideia da análise gerou um leque de informações interessantes e que ocorrem frequentemente no meio do futebol, nem sempre o clube com mais expressão e investimentos ganham todas as competições, lógico, a tendência é que eles sempre chegam nas principais posições mas o desempenho em campo pode mudar os panoramas de expectativas. Em suma, sem organização, empenho e trabalho coletivo não é possível gerar resultados excelentes e esperados para o investimento inserido.

APÊNDICES

Marcone

Código do gráfico de 2015:

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm

# Carregar os dados do arquivo CSV
data = pd.read_csv("2015.csv")

# Extrair as colunas necessárias para o scatter plot
gdp_per_capita = data["Economy (GDP per Capita)"]
perceptions_of_corruption = 300*data["Trust (Government Corruption)"]# alterando tamanho dos pontos ao multiplicar por 300
healthy_life_expectancy = data["Health (Life Expectancy)"]
Freedom = data["Freedom"]

# Definir o tamanho do gráfico
plt.figure(figsize=(12, 10))

# Criar o scatter plot com mapa de cores personalizado
plt.scatter(gdp_per_capita, healthy_life_expectancy, c=Freedom, cmap=cm.RdYlGn, s=perceptions_of_corruption)

# Adicionar rótulos aos eixos
plt.xlabel("PIB per capita")
plt.ylabel("Expectativa de vida saudável")

# Adicionar uma barra de cores para codificar a Generosity
colorbar = plt.colorbar()
colorbar.set_label("Liberdade para fazer escolhas de vida")

# Ajustar a legenda para o tamanho dos pontos (Perceptions of corruption)
plt.legend(title="Percepções de corrupção")

# Exibir o gráfico
plt.show()
```

Código do gráfico de 2021:

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm

# Carregar os dados do arquivo CSV
data = pd.read_csv("2021.csv")

# Extrair as colunas necessárias para o scatter plot
gdp_per_capita = data["Explained by: Log GDP per capita"]
perceptions_of_corruption = 300*data["Explained by: Perceptions of corruption"]# alterando tamanho dos pontos ao multiplicar por 300
healthy_life_expectancy = data["Explained by: Healthy life expectancy"]
Freedom = data["Explained by: Freedom to make life choices"]

# Definir o tamanho do gráfico
plt.figure(figsize=(12, 10))

# Criar o scatter plot com mapa de cores personalizado
plt.scatter(gdp_per_capita, healthy_life_expectancy, c=Freedom, cmap=cm.RdYlGn, s=perceptions_of_corruption)

# Adicionar rótulos aos eixos
plt.xlabel("PIB per capita")
plt.ylabel("Expectativa de vida saudável")

# Adicionar uma barra de cores para codificar a Generosity
colorbar = plt.colorbar()
colorbar.set_label("Liberdade para fazer escolhas de vida")

# Ajustar a legenda para o tamanho dos pontos (Perceptions of corruption)
plt.legend(title="Percepções de corrupção")

# Exibir o gráfico
plt.show()
```

REFERÊNCIAS

World Happiness Report. Disponível em:

<https://www.kaggle.com/search?q=World+Happiness+Report>.