

UNIVERSIDADE DE SÃO PAULO - USP
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO -
ICMC

AFONSO HENRIQUE PIACENTINI GARCIA,
Nº USP: 9795272
BRENO LÍVIO SILVA DE ALMEIDA,
Nº USP: 10276675
VITOR HENRIQUE GRATIERE TORRES,
Nº USP: 10284952

SCC0652 - VISUALIZAÇÃO COMPUTACIONAL
TRABALHO PRÁTICO: VISUALIZANDO O CONJUNTO
DE DADOS DE POKÉMONS

SÃO CARLOS
2020

LISTA DE FIGURAS

Figura 1 – Exemplo posterior à junção.	4
Figura 2 – Nomes de Pokémon com símbolos.	4
Figura 3 – Pokémons com código duplicado.	6
Figura 4 – Pokémon outlier.	6
Figura 5 – Pokémons com valores na extremidade em "hp".	6
Figura 6 – Melhores gerações.	7
Figura 7 – Resumo descritivo dos dados.	7
Figura 8 – Exemplo de palheta padrão.	8
Figura 9 – Exemplo de palheta colorblind.	8
Figura 10 – Gráfico de linha agrupando os Pokémons em gerações para variável total. . .	9
Figura 11 – Boxplot agrupando os Pokémons em gerações para variável total.	9
Figura 12 – Violin plot agrupando os Pokémons em gerações para variável total.	10
Figura 13 – Heatmap para correlação de Pearson.	11
Figura 14 – Heatmap para correlação de Spearman.	11
Figura 15 – Scatter plot entre peso e altura.	12
Figura 16 – Histogramas para as variáveis.	13
Figura 17 – Histograma para peso.	13
Figura 18 – Pair plot para as variáveis quantitativas.	14
Figura 19 – Frequências absolutas dos Pokémons com determinados tipos.	15
Figura 20 – Gráficos de radar para os tipos de Pokémon.	15
Figura 21 – Word cloud lembrando o formato de uma Pokébola.	16
Figura 22 – Interação com hover gerada pelo Plotly e widget.	17
Figura 23 – Zoom na interação com plotly e widget.	18
Figura 24 – Gráfico de peso para altura interativo.	19
Figura 25 – Interação em conjunto de Violin plot com Radar chart.	20
Figura 26 – Junção dos radar charts.	20
Figura 27 – Tipos dos pokémons	21
Figura 28 – Pokédex interativa	22
Figura 29 – Pokédex interativa buscando pelo nome	22
Figura 30 – Radar chart na Pokédex interativa	23
Figura 31 – Lista de Pokémons na Pokédex interativa	23

SUMÁRIO

1	INTRODUÇÃO	3
2	DESENVOLVIMENTO	4
2.1	Processamento de dados	4
2.1.1	Gerando o Data Frame Principal	4
2.1.2	Descrição do conjunto de dados	4
2.1.3	Análise dos dados	5
2.1.3.1	<i>Códigos das Pokédex</i>	5
2.1.3.2	<i>Outliers</i>	6
2.1.3.3	<i>Comparações entre os extremos dos stats</i>	6
2.1.3.4	<i>As melhores gerações</i>	7
2.1.3.5	<i>Resumo descritivo</i>	7
2.2	Visualização de dados	8
2.2.1	Cores das visualizações	8
2.2.2	Visualizando os dados	8
2.2.2.1	<i>Outliers</i>	9
2.2.2.2	<i>Correlações</i>	10
2.2.2.3	<i>Distribuição</i>	12
2.2.2.4	<i>Word cloud</i>	16
2.3	Visualização interativa	17
2.3.1	Cores das visualizações	17
2.3.2	Interagindo com as visualizações dos dados	17
2.3.2.1	<i>Outliers</i>	17
2.3.2.2	<i>Correlação</i>	18
2.3.2.3	<i>Distribuição</i>	19
2.3.3	Pokédex interativa	21
	Conclusão	24
	REFERÊNCIAS	25

1 INTRODUÇÃO

O projeto consiste no processamento, análise e visualização de um conjunto de dados, utilizando diversos conhecimentos e ferramentas obtidos nas aulas da matéria de Visualização Computacional. No caso, foi usado um dataset (conjunto de dados) de Pokémons, do vídeo-game homônimo (POKEMON, 1996).

Para gerar nosso conjunto de dados, buscamos em diversas fontes diferentes, procurando ter o máximo possível de informações relevantes. No fim, parte dos dados foi retirado da Pokemon Database (DATABASE, 2008), e parte de um dataset do Kaggle (KAGGLE, 2020), que utiliza a licença Creative Commons. Por fim, processamos os dados obtidos anteriormente, realizamos uma primeira análise deles, e geramos o nosso próprio conjunto de dados.

Depois disso, realizamos uma primeira visualização dos nossos dados. Para isso, usando alguns dos dados encontrados das análises iniciais e ferramentas computacionais específicas, geramos diversos gráficos para visualizar melhor diferentes correlações dentro do nosso conjunto de informações.

Por fim, usando dados e correlações obtidas anteriormente, e usando ainda mais ferramentas computacionais, conseguimos criar diversas visualizações interativas dessas informações. Com isso, conseguimos, inclusive, gerar nossa própria versão da Pokédex, uma ferramenta presente em todos os jogos da franquia Pokémon, usada pelos treinadores para visualizar informações dos pokémons que esses encontram durante sua jornada.

2 DESENVOLVIMENTO

Vamos trabalhar com os dados dos Pokémons de forma que iremos tratar os dados no início até poder criar visualizações convenientes para o projeto. Cada etapa será explicada de forma concisa e direta comparada ao que foi feito em código. Os códigos podem ser encontrados no repositório do GitHub (REPOSITORIO, 2020).

2.1 Processamento de dados

2.1.1 Gerando o Data Frame Principal

Utilizando do Pandas (PANDAS, 2008), uma biblioteca do Python que serve como uma ferramenta eficiente para análise de dados, carregamos os dados referentes aos Pokémons, misturando dois conjuntos retirados de lugares diferente como anteriormente falado. Os dados são carregados em espécies de tabelas, e foram utilizados alguns conceitos de base de dados para fazer a junção delas.

Figura 1 – Exemplo posterior à junção.

cod	name	generation	status	species	form	type1	type2	height_m	weight_kg	total	hp	attack	defense	spatk	spdef	speed	img
337	284	Masquerain	3	Normal	Eyeball Pokémon	Bug	Flying	0.8	3.6	454	70	60	62	100	82	80	https://img.pokemondb.net/sprites/home/normal/...
345	292	Shedinja	3	Normal	Shed Pokémon	Bug	Ghost	0.8	1.2	236	1	90	45	30	30	40	https://img.pokemondb.net/sprites/home/normal/...
628	536	Palpitoad	5	Normal	Vibration Pokémon	Water	Ground	0.8	17.0	384	75	65	55	65	55	69	https://img.pokemondb.net/sprites/home/normal/...
512	432	Purugly	4	Normal	Tiger Cat Pokémon	Normal		1.0	43.8	452	71	82	64	64	59	112	https://img.pokemondb.net/sprites/home/normal/...
943	813	Scorbunny	8	Normal	Rabbit Pokémon	Fire		0.3	4.5	310	50	71	40	40	40	69	https://img.pokemondb.net/sprites/home/normal/...

Note que além da junção, foram tratados outras situações com linhas que não possuíam valores em colunas específicas, além de renomear Pokémons com símbolos especiais.

Figura 2 – Nomes de Pokémon com símbolos.

cod	name	generation	status	species	form	type1	type2	height_m	weight_kg	total	hp	attack	defense	spatk	spdef	speed	img
40	029	Nidoran♀	1	Normal	Poison Pin Pokémon	Poison		0.4	7.0	275	55	47	52	40	40	41	https://img.pokemondb.net/sprites/home/normal/...
43	032	Nidoran♂	1	Normal	Poison Pin Pokémon	Poison		0.5	9.0	273	46	57	40	40	40	50	https://img.pokemondb.net/sprites/home/normal/...

2.1.2 Descrição do conjunto de dados

Com o conjunto de dados coletado e tratado, finalizamos com as seguintes variáveis representadas por colunas no data frame do Pandas:

- **cod:** O número do pokémon na Pokédex;
- **name:** O nome do pokémon;
- **generation:** a qual geração o pokémon em questão pertence;

- **status:** refere-se a raridade do pokémon, variando de normal até mítico;
- **species:** apelido dado para cada pokémon baseado em algum de seus atributos;
- **form:** discriminante de mega evoluções e outras formas;
- **type1:** tipo do pokémon;
- **type2:** segundo tipo do pokémon, caso possua;
- **height_m:** altura do pokémon em metros;
- **weight_kg:** peso do pokémon em kilogramas.
- **total:** soma total dos stats do pokémon;
- **hp:** stat que indica a vida base do pokémon;
- **attack:** stat que indica o poder de ataque físico base do pokémon;
- **defense:** stat que indica a defesa física base do pokémon;
- **spatk:** stat que indica o poder de ataque especial do pokémon;
- **spdef:** stat que indica o poder de defesa especial do pokémon;
- **speed:** stat que indica a velocidade base do pokémon;
- **img:** link para a imagem do pokémon;

2.1.3 Análise dos dados

Com o pré-processamento dos dados, foram feitas algumas análises relevantes ao conjunto de dados que possuímos, principalmente para visualizar e entender melhor nas etapas seguintes.

2.1.3.1 Códigos das Pokédex

Podemos analisar os códigos duplicados dos Pokémons para tentar entender o comportamento dessa situação.

Isso ocorre por conta da forma que a Pokédex usa o mesmo código para lidar com formas alternativas dessas criaturas.

Figura 3 – Pokémons com código duplicado.

	cod	name	generation	status	species	form	type1	type2	height_m	weight_kg	total	hp	attack	defense	spatk	spdef	speed	img	
	2	003	Venusaur	1	Normal	Seed Pokémon	Grass	Poison	2.0	100.0	525	80	82	83	100	100	80	https://img.pokemondb.net/sprites/home/normal/...	
	3	003	Venusaur	1	Normal	Seed Pokémon	Mega Venusaur	Grass	Poison	2.4	155.5	625	80	100	123	122	120	80	https://img.pokemondb.net/sprites/home/normal/...
	6	006	Charizard	1	Normal	Flame Pokémon	Fire	Flying	1.7	90.5	534	78	84	78	109	85	100	https://img.pokemondb.net/sprites/home/normal/...	
	7	006	Charizard	1	Normal	Flame Pokémon	Mega Charizard X	Fire	Dragon	1.7	110.5	634	78	130	111	130	85	100	https://img.pokemondb.net/sprites/home/normal/...
	8	006	Charizard	1	Normal	Flame Pokémon	Mega Charizard Y	Fire	Flying	1.7	100.5	634	78	104	78	159	115	100	https://img.pokemondb.net/sprites/home/normal/...
...	
1023	888	Zacian	8	Legendary	Warrior Pokémon	Hero of Many Battles	Fairy		2.8	110.0	670	92	130	115	80	115	138	https://img.pokemondb.net/sprites/home/normal/...	
1024	889	Zamazenta	8	Legendary	Warrior Pokémon	Crowned Shield	Fighting	Steel	2.9	785.0	720	92	130	145	80	145	128	https://img.pokemondb.net/sprites/home/normal/...	
1025	889	Zamazenta	8	Legendary	Warrior Pokémon	Hero of Many Battles	Fighting		2.9	210.0	670	92	130	115	80	115	138	https://img.pokemondb.net/sprites/home/normal/...	
1026	890	Eternatus	8	Legendary	Gigantic Pokémon		Poison	Dragon	20.0	950.0	690	140	85	95	145	95	130	https://img.pokemondb.net/sprites/home/normal/...	
1027	890	Eternatus	8	Legendary	Gigantic Pokémon	Eternamax	Poison	Dragon	100.0	NaN	1125	255	115	250	125	250	130	https://img.pokemondb.net/sprites/home/normal/...	

2.1.3.2 Outliers

Utilizando o conceito de distância interquartil, podemos encontrar Pokémons que possuem valores discrepantes comparado ao resto do conjunto. O cálculo é feito da forma a seguir:

$$\text{outliers} = x_i < (Q_1 - k.IQR) \vee x_i > (Q_3 + k.IQR)$$

Foi usado $k = 1,5$, como geralmente usam para detectar outliers em datasets mais gerais. Então podemos ver os Pokémons que são outliers com base na variável "total":

Figura 4 – Pokémon outlier.

cod	name	generation	status	species	form	type1	type2	height_m	weight_kg	total	hp	attack	defense	spatk	spdef	speed	img	
1027	890	Eternatus	8	Legendary	Gigantic Pokémon	Eternamax	Poison	Dragon	100.0	NaN	1125	255	115	250	125	250	130	https://img.pokemondb.net/sprites/home/normal/...

Vamos explorar melhor esse conceito na parte de visualização dos dados.

2.1.3.3 Comparações entre os extremos dos stats

Vamos também explorar os Pokémons que se encontram nas extremidades dos valores de stats para eles. O próprio Pandas tem a função adequada para possibilitar isso. Podemos fazer, por exemplo, para a variável "hp":

Figura 5 – Pokémons com valores na extremidade em "hp".

cod	name	generation	status	species	form	type1	type2	height_m	weight_kg	total	hp	attack	defense	spatk	spdef	speed	img
288	242	Blissey	2	Normal	Happiness Pokémon	Normal		1.5	46.8	540	255	10	10	75	135	55	https://img.pokemondb.net/sprites/home/normal/...



cod	name	generation	status	species	form	type1	type2	height_m	weight_kg	total	hp	attack	defense	spatk	spdef	speed	img
345	292	Shedinja	3	Normal	Shed Pokémon	Bug	Ghost	0.8	1.2	236	1	90	45	30	30	40	https://img.pokemondb.net/sprites/home/normal/...



Foi também feito de forma análoga para os outros stats dos Pokémons.

2.1.3.4 As melhores gerações

Podemos encontrar a melhor geração de Pokémons baseado na média dos valores da variável "total" dos Pokémons de cada geração. Para isso foi necessária a função de agrupamento do Pandas para agrupar os Pokémons em suas gerações e assim tirar a média para cada uma.

Figura 6 – Melhores gerações.

	total	hp	attack	defense	spatk	spdef	speed
generation							
7	459.232323	71.464646	87.313131	79.202020	77.848485	75.555556	67.848485
4	459.016529	73.082645	82.867769	78.132231	76.404959	77.190083	71.338843
6	442.082353	69.882353	77.070588	76.905882	75.141176	75.082353	68.000000
8	438.329545	70.465909	80.000000	75.125000	71.772727	72.454545	68.511364
3	436.006061	66.806061	81.030303	74.048485	75.600000	71.139394	67.381818
5	435.157895	72.292398	82.976608	72.269006	70.929825	68.426901	68.263158
1	424.104167	64.760417	77.302083	71.005208	69.927083	68.598958	72.510417
2	419.140187	71.476636	71.869159	73.822430	66.121495	74.336449	61.514019

Com as médias das gerações, colocamos em ordem decrescente, indicando que a melhor geração é a primeira linha da tabela obtida.

2.1.3.5 Resumo descritivo

Podemos fazer o resumo das medidas estatísticas dos dados por meio de uma função do Pandas.

Figura 7 – Resumo descritivo dos dados.

	generation	height_m	weight_kg	total	hp	attack	defense	spatk	spdef	speed
count	1028.000000	1028.000000	1027.000000	1028.000000	1028.000000	1028.000000	1028.000000	1028.000000	1028.000000	1028.000000
mean	4.034047	1.368093	69.753749	437.533074	69.577821	80.109922	74.465953	72.722763	72.122568	68.534047
std	2.234937	3.380126	129.221230	121.641715	26.385849	32.352799	31.281363	32.656156	28.058406	29.802103
min	1.000000	0.100000	0.100000	175.000000	1.000000	5.000000	5.000000	10.000000	20.000000	5.000000
25%	2.000000	0.600000	8.800000	330.000000	50.000000	55.000000	50.000000	50.000000	50.000000	45.000000
50%	4.000000	1.000000	28.500000	455.000000	66.500000	76.000000	70.000000	65.000000	70.000000	65.000000
75%	6.000000	1.500000	69.100000	510.000000	80.000000	100.000000	90.000000	95.000000	90.000000	90.000000
max	8.000000	100.000000	999.900000	1125.000000	255.000000	190.000000	250.000000	194.000000	250.000000	180.000000

2.2 Visualização de dados

Com o carregamento de dados feito, conseguimos iniciar o processo de visualização de dados.

2.2.1 Cores das visualizações

Antes de propriamente visualizar os dados, é importante ressaltar a escolha de uma palheta adequada de cores. Na biblioteca do Seaborn, temos diversas opções de palhetas de cores, e iremos trabalhar com a opção "colorblind" para tentar tornar as visualizações mais acessíveis.

Figura 8 – Exemplo de palheta padrão.



Figura 9 – Exemplo de palheta colorblind.



Entretanto, algumas situações específicas, como as cores dos tipos dos Pokémons e a Pokébola feita em Word cloud, seguirão os padrões esperados da franquia de videogame. Algumas cores em questão servem para tornar a visualização mais fiel ao conjunto de dados.

2.2.2 Visualizando os dados

A ideia da visualização de dados é muito importante para possibilitar uma identificação mais rápida e precisa sobre informações relevantes de um conjunto de dados, fazendo com que até mesmo leigos sobre um assunto possam pelo menos ter uma breve noção do comportamento dos dados em questão. Para isso precisamos fazer visualizações realmente pensadas e justificadas com base no tipo de dados que estamos lidando.

Esta seção irá servir não somente para mostrar algumas informações interessantes (como outliers, relações, entre outros) por meio de vários tipos de visualizações, mas também para justificar o uso dessas.

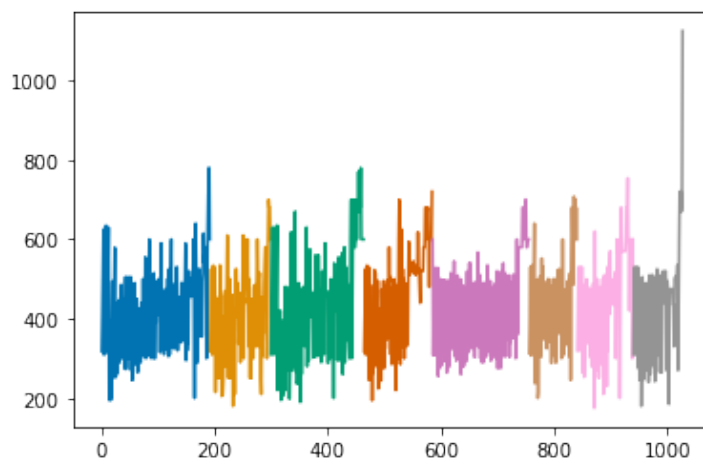
Vale lembrar que estamos lidando com um dataset híbrido, contendo diversas categorias como generation, status e type1, e também contendo diversas variáveis numéricas como height_m, weight_kg, total, hp, attack, defense, entre outros. O catálogo From Data to Viz (FROM-DATA-TO-VIZ, 2018) nos mostra diversas visualizações que podemos fazer com base nesse conjunto de dados híbrido.

2.2.2.1 Outliers

Podemos encontrar Outliers por meio de diferentes tipos de visualizações. Algumas possíveis formas são demonstradas a seguir, escolhidas baseadas na sua relevância para os dados analisados no projeto.

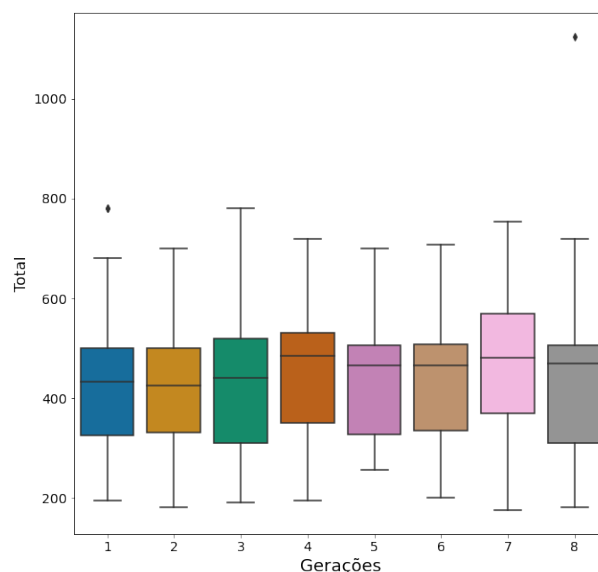
Podemos observar outliers bastando apenas fazer gráficos de linha. Então, por exemplo, se quisermos detectar outliers visualmente, é possível notar uma discrepância da forma que um ponto acaba levando uma linha.

Figura 10 – Gráfico de linha agrupando os Pokémons em gerações para variável total.



Mas o mais adequado realmente para achar outliers talvez seja o boxplot, que trabalha com a ideia do intervalo interquartil e ver o quanto os dados escapam desse intervalo. Dessa forma, no gráfico de linha, notamos o mesmo outlier novamente na oitava geração.

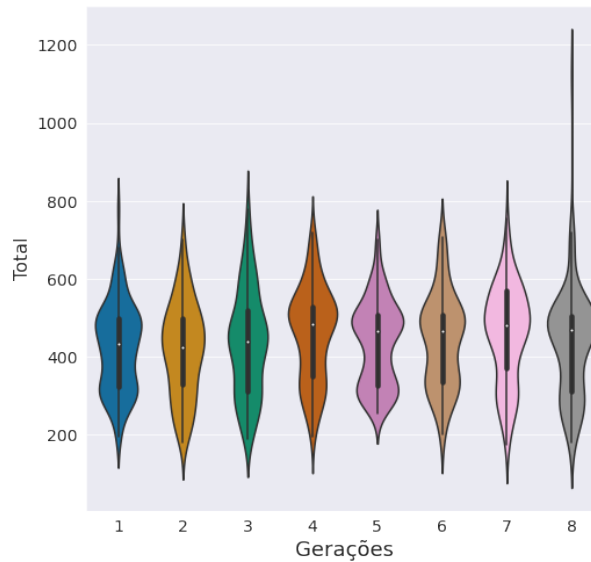
Figura 11 – Boxplot agrupando os Pokémons em gerações para variável total.



Outra alternativa ao gráfico boxplot é o gráfico de violino. Tal visualização permite a criação de um gráfico intuitivo no qual pode-se ver onde os dados se acumulam e onde eles

são menos frequentes, mostrando a distribuição dos dados ao longo da reta. A visualização do boxplot junto à de violino, permite uma boa análise visual do comportamento dos dados, onde sua mediana se encontra e onde os outliers estão localizados. Segue abaixo gráficos de violino para as mesmas variáveis estudadas anteriormente.

Figura 12 – Violin plot agrupando os Pokémons em gerações para variável total.



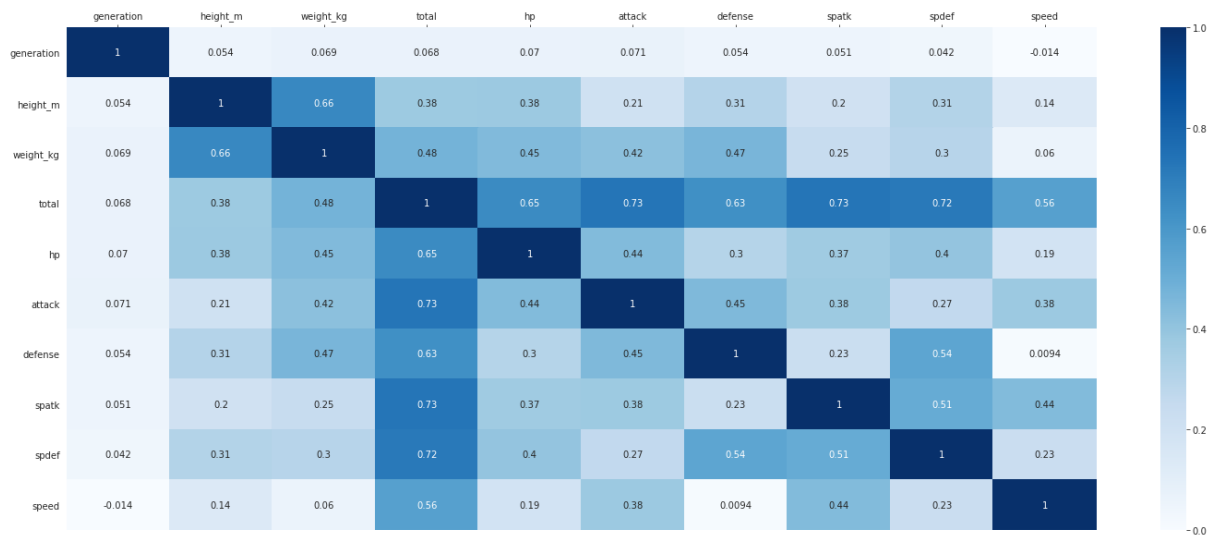
2.2.2.2 Correlações

Encontrar correlações em um conjunto de dados pode ser sempre interessante para descobrir algumas possíveis ligações entre variáveis, as quais nem sempre eram esperadas. Vamos visualizar essas correlações por meio de Heatmaps, usando as conhecidas correlações de Pearson e Spearman.

Antes de tudo, é aconselhável a normalização dos dados para se fazer visualizações com Heatmap. Entretanto, como apenas queremos fazer Heatmaps para ver as correlações de Pearson e Spearman, sendo estas já normalizadas por definição, podemos fazer diretamente sem perda de dados. O coeficiente da correlação de Pearson para uma amostra pode ser calculado com seguinte:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

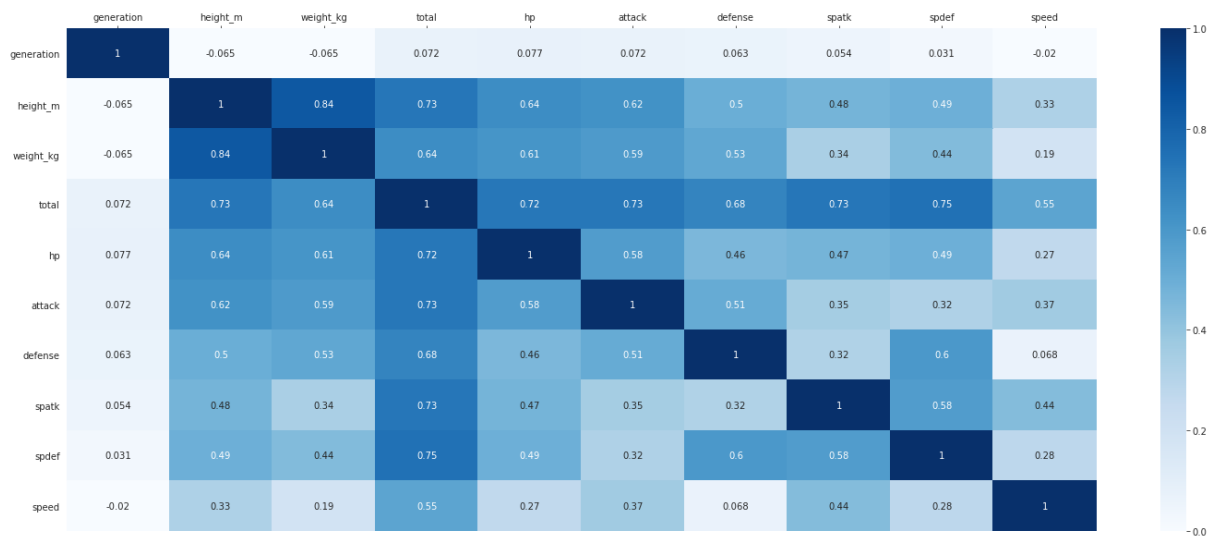
Figura 13 – Heatmap para correlação de Pearson.



Já para o coeficiente da correlação de Spearman temos:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Figura 14 – Heatmap para correlação de Spearman.



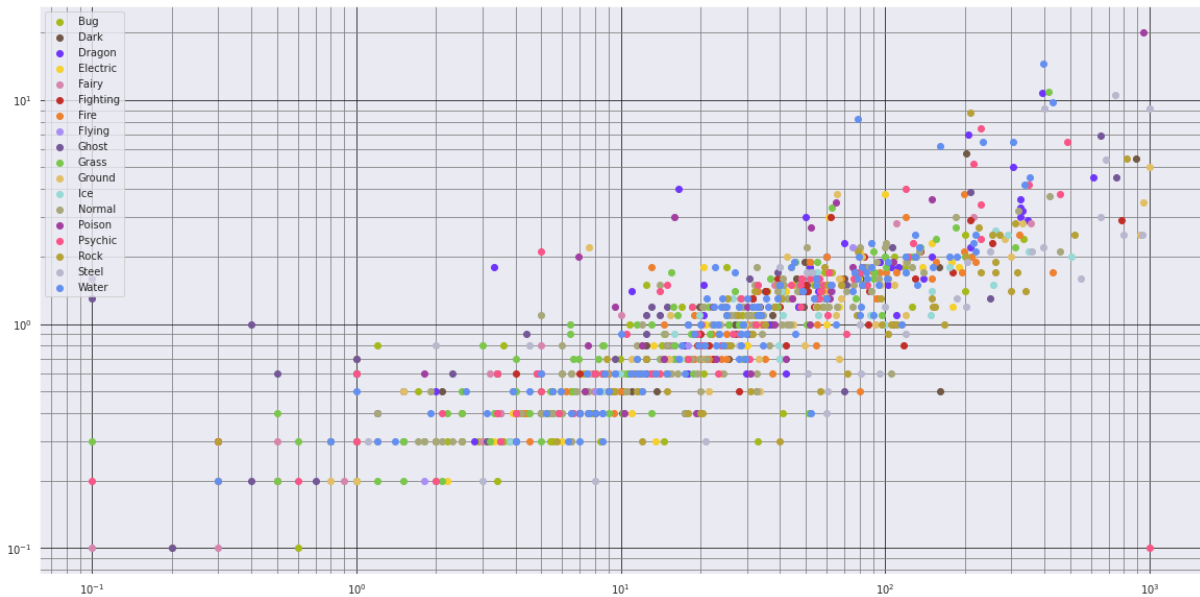
É interessante observar que em Spearman as correlações se mostraram mais fortes que no caso da correlação de Pearson. Isso ocorre porque os dados que estamos tratando não são dispostos de uma forma linear, o que prejudica o cálculo do coeficiente da correlação de Pearson. Dessa forma, o Spearman se mostrou mais efetivo.

Algumas correlações relevantes com certeza foram as esperadas para os stats dos Pokémon, pois total representa a soma de todos os stats hp, attack, defense, spatk, spdef, speed, então faz sentido ter uma correlação um tanto mais elevada.

A correlação não tão esperada e ampliada por Spearman, foi sobre a altura (height_m) e o peso (weight_kg) dos Pokémons. Mostrando uma certa tendência de quanto mais alto for o Pokémon mais pesado ele será.

Dado isso, vamos fazer um Scatter plot com os valores em logaritmo para vermos a relação de maneira mais clara.

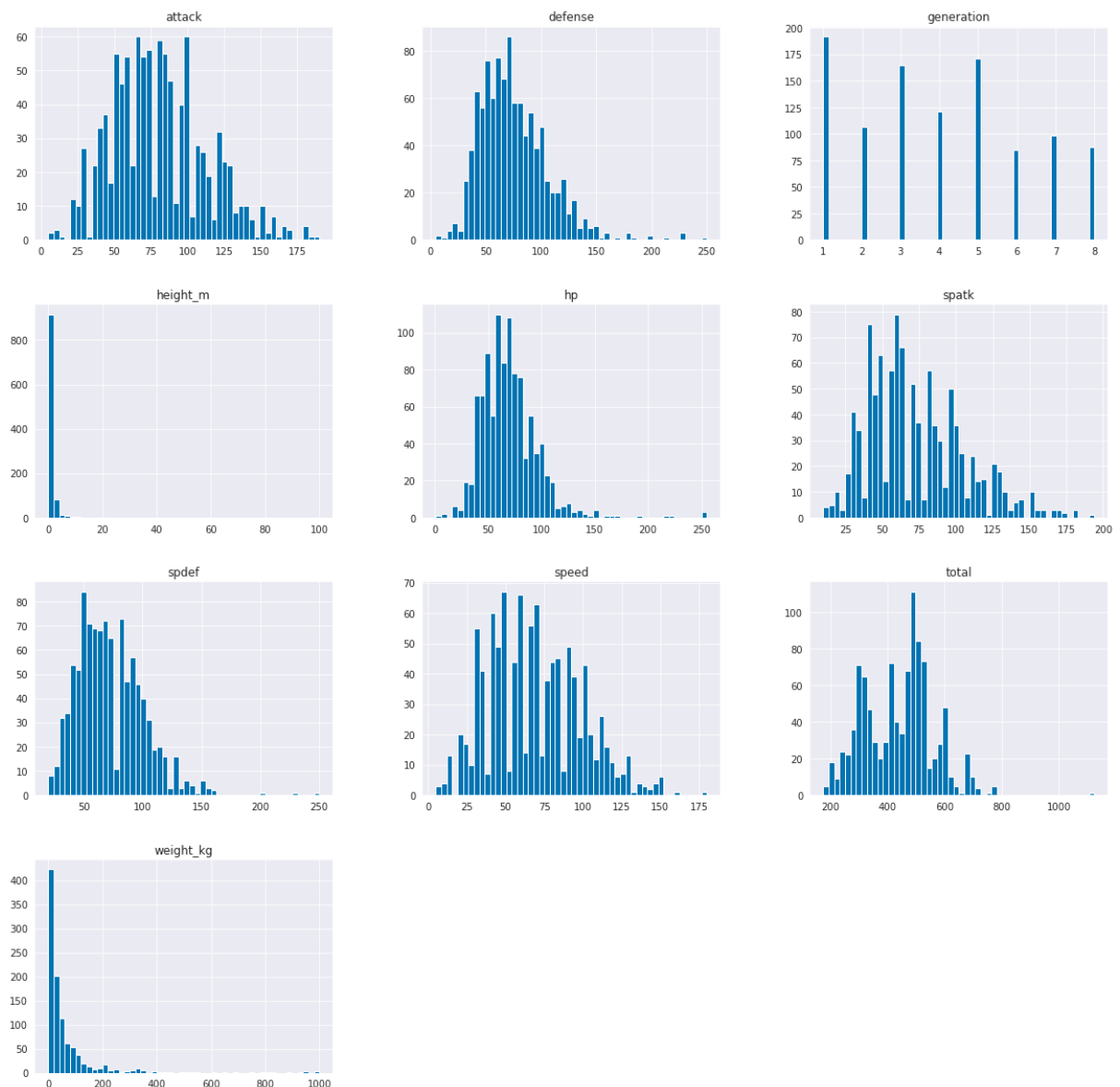
Figura 15 – Scatter plot entre peso e altura.



2.2.2.3 Distribuição

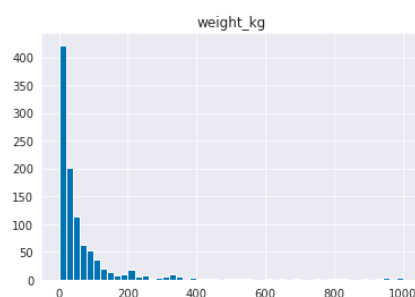
Vamos visualizar a distribuição das variáveis cabíveis por meio de histogramas. Devemos acrescentar que um histograma realmente é um gráfico de barras, entretanto as observações feitas em bins devem ter área igual a 1. Além disso, é interessante ter um número considerável de bins para não haver perda de dados.

Figura 16 – Histogramas para as variáveis.



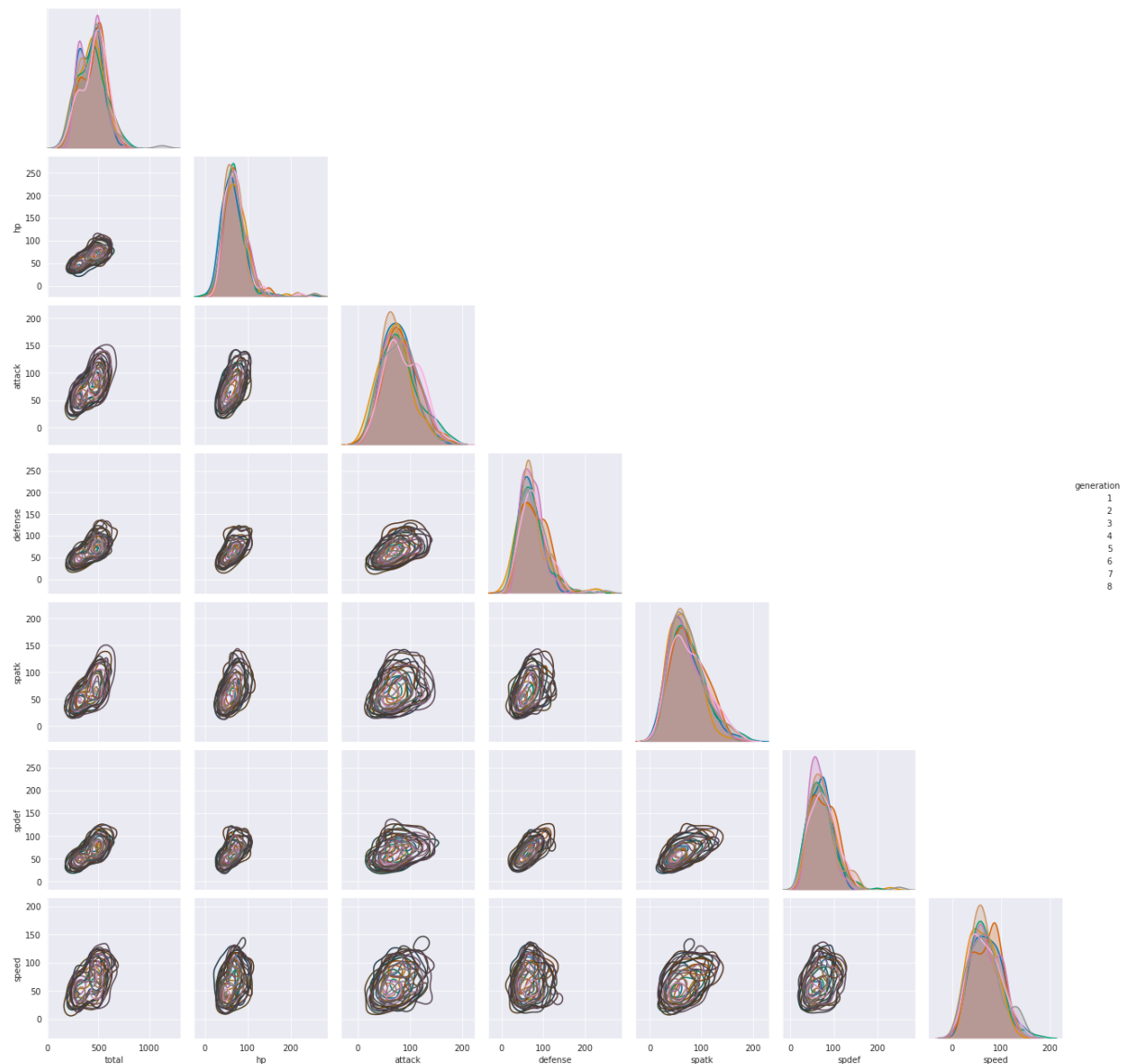
Podemos notar que algumas distribuições podem se parecer com alguma distribuição mais específica, como uma Gaussiana, mas isso nem sempre se verifica. Entretanto podemos fazer alguns teste como o de Shapiro-Wilk para testar a hipótese na visualização feita. No caso, temos um exemplo que funciona para a distribuição da variável `weight_kg`.

Figura 17 – Histograma para peso.



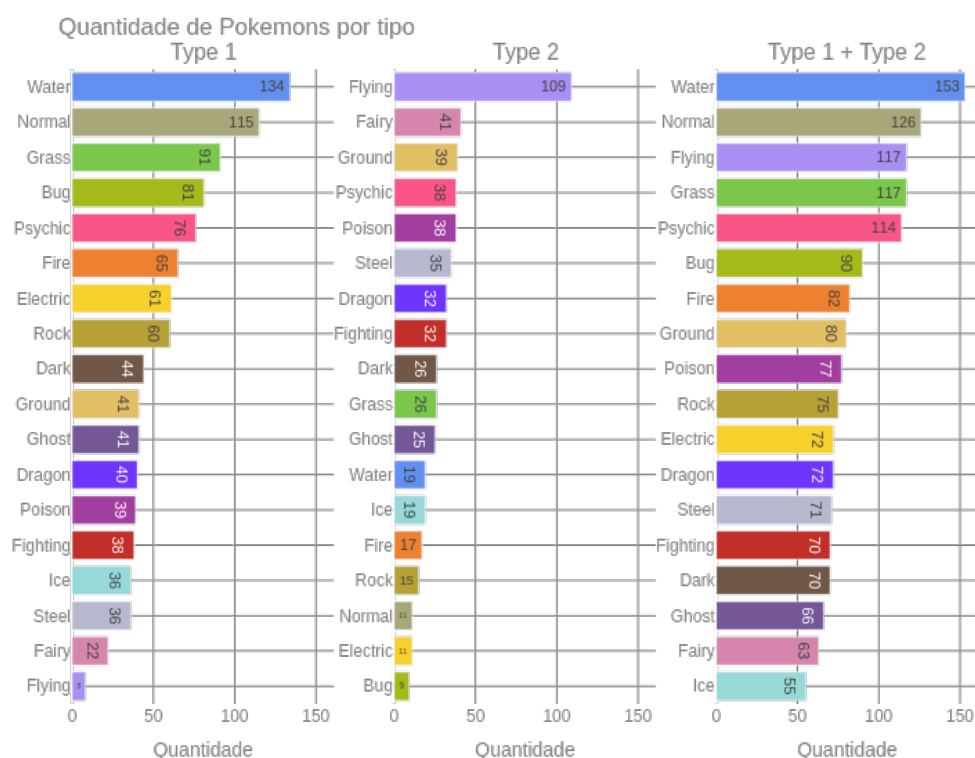
Outra forma interessante de analisar distribuições de variáveis seria o uso do Pair Plot. Nele, podemos observar várias distribuições agrupadas de forma que nos permita visualizar relações entre as diferentes variáveis no nosso conjunto de dados. Na distribuição a seguir veremos distribuições agrupadas por gerações analisando as variáveis possíveis como total, hp, entre outras.

Figura 18 – Pair plot para as variáveis quantitativas.



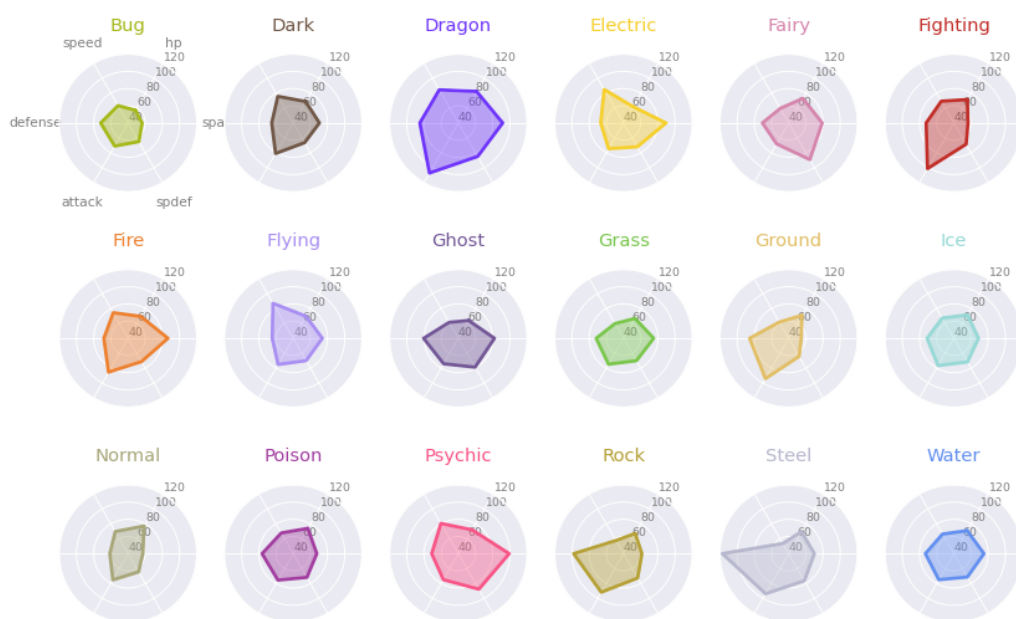
Outro dado interessante que podemos visualizar a distribuição é a da quantidade de tipos de Pokémons. Podemos trabalhar com um simples gráfico de barras, analisando a frequência absoluta de cada tipo. Como cada Pokémon pode ter dois tipos com Type 1 e Type 2, iremos ver a distribuição para cada um e a soma de ambos.

Figura 19 – Frequências absolutas dos Pokémons com determinados tipos.



Também podemos visualizar esses tipos em questão para entender exatamente como são os Pokémons que pertencem a essas categorias. Dessa forma, vamos ver a média dos valores dos stats hp, attack, defense, speed, spatk e spdef por meio de um Radar chart. Ele vai nos permitir facilmente entender para qual stat um Pokémon de tal tipo tende a ser melhor, por exemplo.

Figura 20 – Gráficos de radar para os tipos de Pokémon.



2.2.2.4 Word cloud

Outra visualização interessante e famosa a ser feita é a nuvem de palavras (word cloud). Por meio dela podemos gerar uma visualização agradável, mesmo para os leigos em visualizações de dados. Vamos fazer uma nuvem de palavras usando da imagem que lembra uma Pokébola do vídeo-game, fazendo com que cada palavra seja um Pokémon e o tamanho de sua palavra seja relacionado com a variável que representa a soma de seus stats, a total. Como estamos usando a Pokébola com uso acadêmico, entra na ideia de "Fair use", e não há problemas com direitos de imagem.

Figura 21 – Word cloud lembrando o formato de uma Pokébola.



2.3 Visualização interativa

Com o carregamento dos dados feito, conseguimos iniciar o processo de visualização interativa dos dados.

2.3.1 Cores das visualizações

Novamente, como na etapa anterior, iremos usar cores específicas que se associam aos tipos de Pokémons que temos no conjunto de dados.

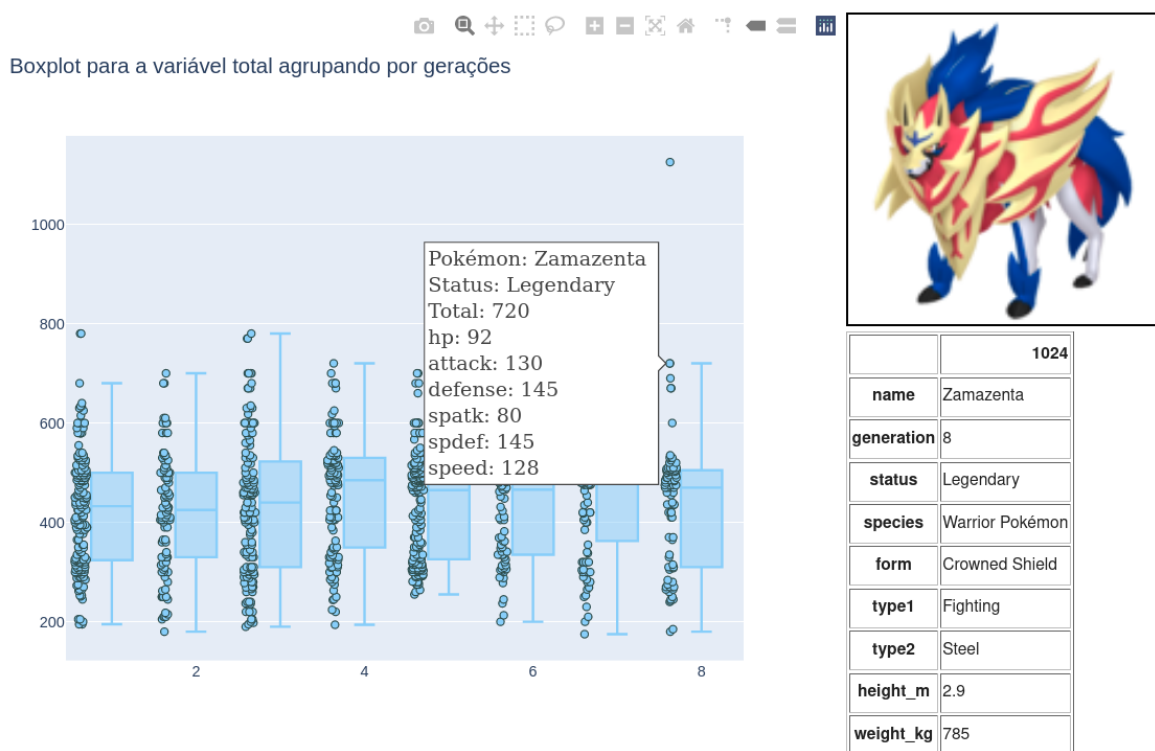
2.3.2 Interagindo com as visualizações dos dados

Vamos desenvolver vários tipos de visualizações interativas, inclusive revisitando visualizações feitas no projeto anterior. Dessa forma espera-se criar um sistema de visualização analítica que nos permita explorar e entender mais ainda sobre nossos dados.

2.3.2.1 Outliers

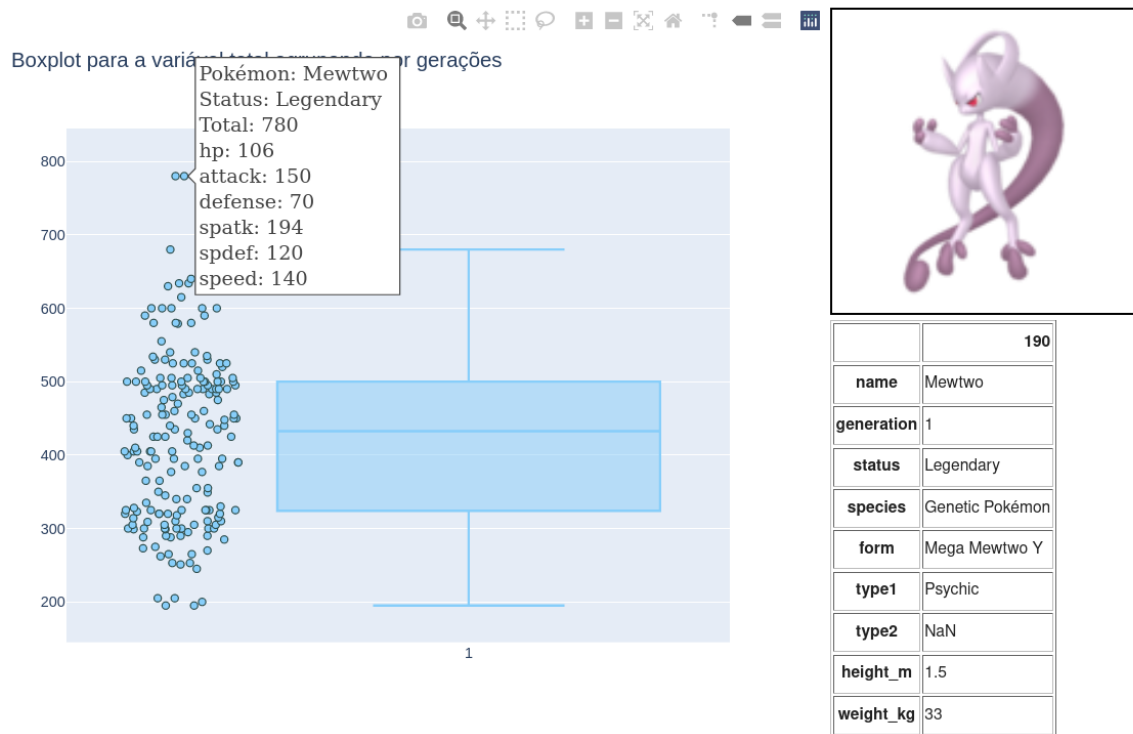
Vamos utilizar a mesma ideia de possivelmente encontrar outliers usando de Boxplot. Dessa vez iremos usar a versão interativa do Plotly junto com a ideia de widget que nos permite visualizar até mesmo quem seria os outliers. Assim basta para usar do hover para atualizar a foto e dados do Pokémon na direita.

Figura 22 – Interação com hover gerada pelo Plotly e widget.



Podemos até perceber casos anteriormente que na visualização estática do boxplot alguns dados eram mais dificilmente percebidos, como a presença de mais de um outlier com o mesmo valor como no caso a seguir.

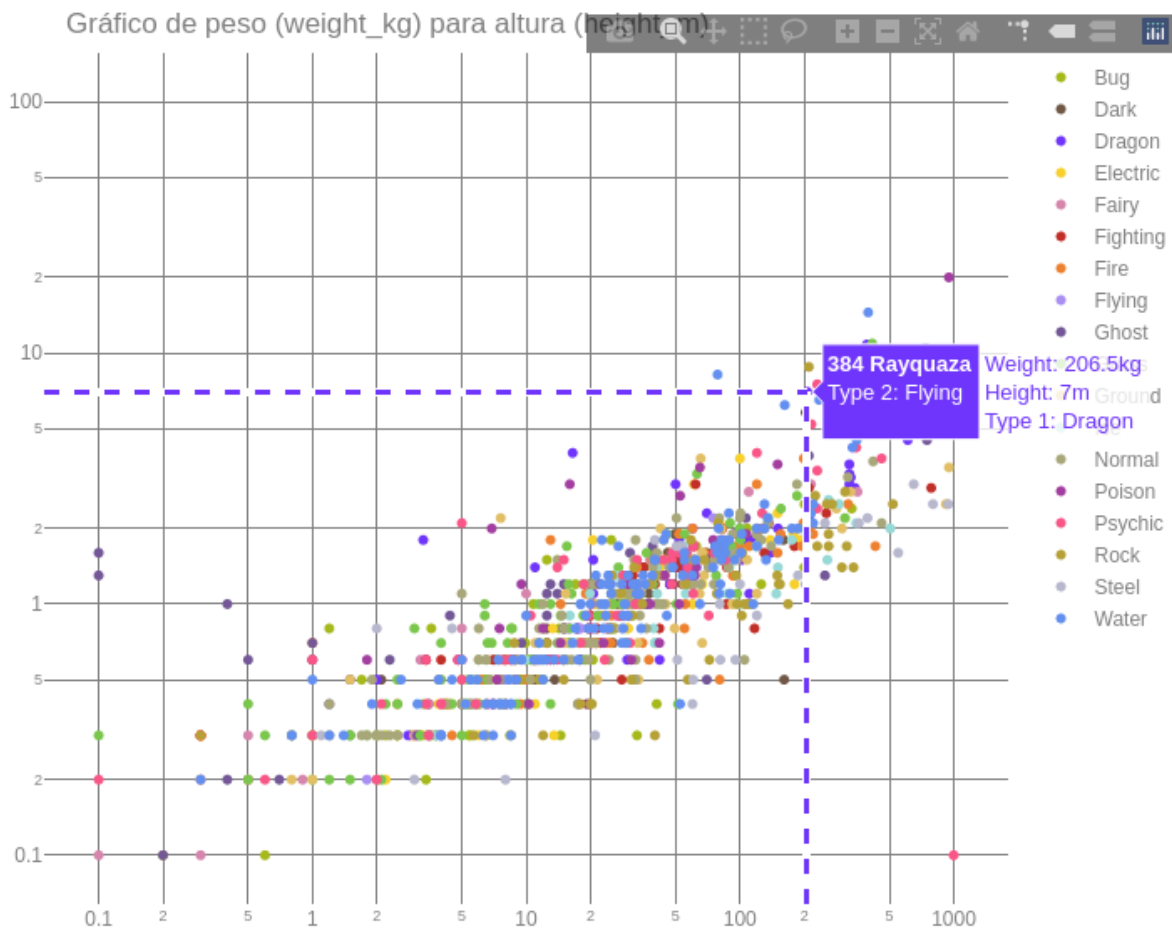
Figura 23 – Zoom na interação com plotly e widget.



2.3.2.2 Correlação

Anteriormente, havíamos criado um Scatter plot analítico para enxergar melhor uma correlação pouco esperada nos dados analisados, no caso a relação entre o peso e a altura dos Pokémon, que no geral se mostrou mutualmente crescente. Dessa vez, decidimos recriar este gráfico em sua forma interativa, na qual além de ser possível visualizar a correlação explicada anteriormente, podem-se checar informações específicas de qualquer um dos pontos no gráfico. No caso, as informações que podem ser checadas são o nome do Pokémon, seu id (ou código), a forma do Pokémon, seus tipos, sua altura e seu peso.

Figura 24 – Gráfico de peso para altura interativo.



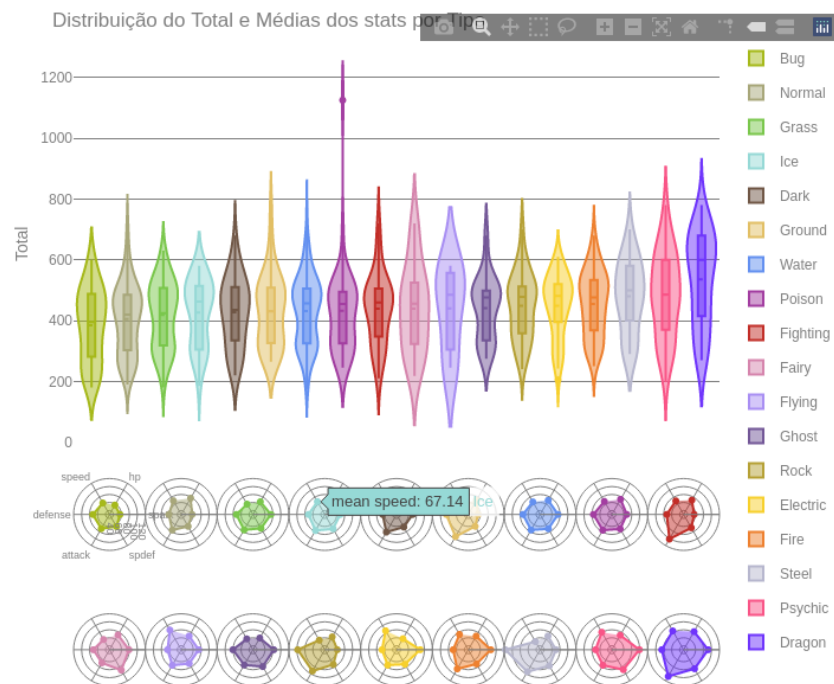
2.3.2.3 Distribuição

Anteriormente, nós geramos diversos radar plots dos stats de cada tipo de Pokémon e violin plots da soma total dos stats dos Pokémon por geração, todos eles de forma estática. Dessa vez, percebemos que tais gráficos poderiam ser muito úteis para o projeto se realizados novamente, porém dessa vez de forma interativa.

Primeiramente, fizemos um violin plot interativo para cada tipo de Pokémon - contendo dentro de cada um o box plot equivalente - nos quais é possível checar diversos dados interessantes de cada violin plot, como o valor mínimo, o máximo, a média e a mediana.

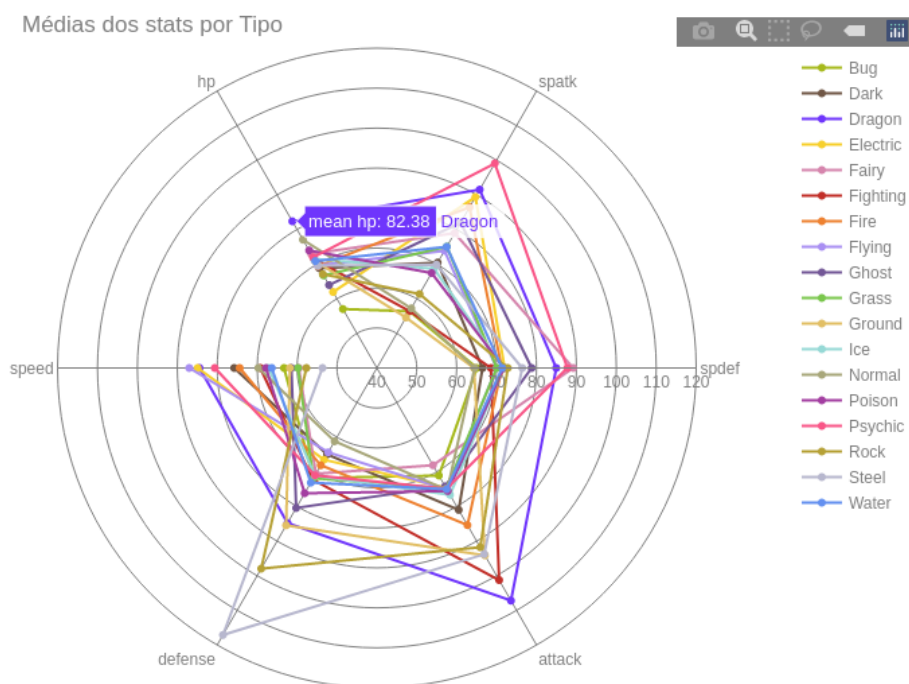
Em segundo lugar, fizemos um radar chart para cada tipo de Pokémon, contendo a média dos stats dos Pokémon daquele tipo específico.

Figura 25 – Interação em conjunto de Violin plot com Radar chart.



Após isso, decidimos agrupar todos os radar charts realizados anteriormente em um único spider chart, tornando a visualização da comparação entre a média dos stats de cada tipo de pokémon ainda mais clara.

Figura 26 – Junção dos radar charts.



2.3.3 Pokédex interativa

Na série de jogos Pokémon, a Pokédex é uma ferramenta indispensável na jornada de qualquer treinador Pokémon, expondo várias informações relevantes sobre qualquer Pokémon que esse treinador já tenha encontrado durante a sua jornada.

Baseando-se nas funcionalidades da Pokédex dos jogos, criamos a nossa versão, que apesar de mais simples que a original, cumpre muito bem a função de servir como, provavelmente, a melhor forma de se visualizar os dados obtidos pelo nosso dataset de forma interativa. Nela, os usuários podem pesquisar por um Pokémon usando o seu id na pokédex original, ou escolhê-lo de uma lista sequencial, que sempre se inicia no pokémon de id 1, Bulbasaur. A Pokédex irá exibir o nome do pokémon, sua imagem, seu tipo (ou seus tipos, se tiver mais de um), seu peso, sua altura, seu id e sua espécie, além de um gráfico de radar com todos os seus stats.

Para a próxima interação gostaríamos de exibir os tipos pokémons com as suas imagens originais, então achamos esse arquivo com todos os tipos de pokémons, porém precisamos cortá-lo, o que é feito por código.

Figura 27 – Tipos dos pokémons



Por fim, iremos fazer a Pokédex funcionando em si, permitindo a busca dos pokémons, visualização dos stats, entre outros. Ela é construída a partir dos widgets do Jupyter que permitem juntar gráficos, criar botões que contém eventos, podendo realmente gerar um sistema de interação para o usuário visualizar informações da maneira que bem desejar.

Figura 28 – Pokédex interativa



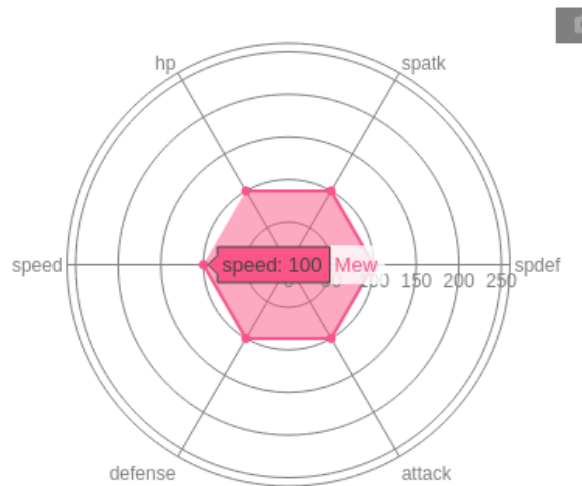
Nela o usuário pode procurar o Pokémon tanto pelo seu código da Pokédex, como mostrado acima, como por seu nome, como mostrado no exemplo abaixo com o pokémon Pikachu.

Figura 29 – Pokédex interativa buscando pelo nome



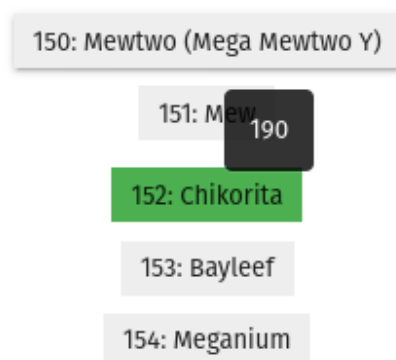
Na Pokédex contamos com a visualização dos stats do Pokémon escolhido por meio de um radar chart já mostrado anteriormente. De forma análoga como mostrado antes, os stats são representados em seis divisões.

Figura 30 – Radar chart na Pokédex interativa



Quando um Pokémon é selecionado aparece uma lista dos Pokémons mais próximos a ele em uma lista para que o usuário pode clicar e visualizar.

Figura 31 – Lista de Pokémons na Pokédex interativa



Observação: Geralmente quando executado de primeira vez parece que ocorre alguns erros de execução, basta interagir buscando um Pokémon no campo de busca que já é resolvido.

CONCLUSÃO

A visualização computacional é muito importante para fazer com que mesmo aqueles leigos a certo assunto possam ter certo entendimento dele por conta da inferência criada a cerca das visualizações. Assim, realizar o projeto foi uma experiência muito próspera. Isso porque, durante o desenvolvimento do trabalho, tivemos que aprender diversos conceitos e alcançar diferentes resultados, todos eles responsáveis por tornar mais simples e direta a visualização de um conjunto de dados.

Além disso, vale dizer que, ao menos em visualização computacional, o conjunto de dados a ser analisado tem grande importância na forma como esse deve ser tratado. No caso, os dados dos pokémons, da franquia de jogos homônima, serviu muito bem para nossos propósitos, já que esses podem gerar visualizações muito interessantes e bem diferentes, a exemplo da Pokédex que conseguimos recriar com certa semelhança para a original.

A Pokédex acabou servindo como a conclusão imprescindível para juntar os conhecimentos adquiridos na parte prática da disciplina, passando desde a parte de processamento de dados, pela visualização desses dados, e finalmente, pela interação com eles. A Pokédex conseguiu reunir toda a ideia do projeto como um todo, permitindo que até mesmo pessoas que não conheçam muito bem o jogo, consigam ao menos ter uma breve noção pelas visualizações e interações criadas.

REFERÊNCIAS

DATABASE. 2008. Site do Pokémon database. Disponível em: <<https://pokemondb.net/pokedex/all>>. Citado na página 3.

FROM-DATA-TO-VIZ. 2018. Site do From Data to Viz. Disponível em: <<https://www.data-to-viz.com/>>. Citado na página 8.

KAGGLE. 2020. Dataset do Kaggle. Disponível em: <<https://www.kaggle.com/mariotormo/complete-pokemon-dataset-updated-090420>>. Citado na página 3.

PANDAS. 2008. Biblioteca Pandas. Disponível em: <<https://pandas.pydata.org/>>. Citado na página 4.

POKEMON. 1996. Site do jogo Pokémon. Disponível em: <<https://www.pokemon.com/br/>>. Citado na página 3.

REPOSITARIO. 2020. Repositório no GitHub. Disponível em: <https://github.com/brenoslivio/SCC0652_Computational_Visualization>. Citado na página 4.