

**Disciplina:** Métodos Quantitativos Aplicados à Administração de Empresas I

**Código da Turma:** 2017202

**Docente:** Profa. Dra. Ana Amélia Benedito Silva

**Grupo 8:** Lauro Hiroshi Pimentel Masuda - 9875437  
Silas Fernandes Moreira – 9761718  
Victor Gomes de Oliveira Martins Nicola - 9844881

# Relatório MQA:

## Pokémon with stats

São Paulo,  
Outubro de 2017

## **Introdução:**

Pokémon é uma série de mídias criada pela empresa *Game Freak*, idealizada pelo japonês Satoshi Tajiri e produzida pelo diretor de arte, Ken Sugimori. A franquia, atualmente, pertence à The Pokémon Company. A franquia teve início com uma série de jogos para vídeo game e hoje conta com série animada, filmes, mangás e vários jogos para outras plataformas.

O trabalho foi centrado em dados sobre atributos de criaturas existentes nos jogos da série original. O objetivo desses jogos é capturar essas criaturas - os Pokémon - e treiná-las para melhorar seus atributos a fim de batalhar com criaturas de outras pessoas, sejam elas treinadores virtuais (parte do jogo) ou jogadores reais (outras pessoas com o jogo).

Além dos atributos de batalha, os Pokémon são divididos por outras características que serão melhor detalhadas. Ao longo do texto buscar-se-á estabelecer relações entre as características e extrair informações mais detalhadas acerca do universo Pokémon.

### **1. Banco de dados:**

O banco de dados escolhido (Pokémon with stats) foi retirado do site *Kaggle* (<https://www.kaggle.com/abcsds/Pokémon>) ele contém status que são usados em uma batalha, abrangendo até a 6ª geração de Pokémon, até o término deste trabalho há 7 gerações de Pokémon. As informações contidas nele foram baseadas nos site: *bulbapedia* ([https://bulbapedia.bulbagarden.net/wiki/Main\\_Page](https://bulbapedia.bulbagarden.net/wiki/Main_Page)), *Pokémon Database* (<https://Pokémond.net/>) e no site oficial do Pokémon (<https://www.Pokémon.com/us/>). A Figura 1 mostra como estão dispostas as informações do banco.

#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
1	Bulbasaur	Grass	Poison	318	45	49	49	65	65	45	1	FALSE
2	Ivysaur	Grass	Poison	405	60	62	63	80	80	80	1	FALSE
3	Venusaur	Grass	Poison	525	80	82	83	100	100	80	1	FALSE
3	VenusaurMega Venu	Grass	Poison	625	80	100	123	122	120	80	1	FALSE
4	Charmander	Fire		309	39	52	43	60	50	65	1	FALSE
5	Chameleon	Fire		405	58	64	58	80	65	80	1	FALSE

Figura 1: recorte do banco de dados

A tabela possui um total de 800 ocorrências, mas há um total de 721 espécies Pokémon. Essa discrepância ocorre devido às variações de forma em algumas espécies, ou seja, um mesmo Pokémon pode ter uma forma especial e seus atributos serão diferentes do original, como exemplificado na Figura 2.

#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk
1	Bulbasaur	Grass	Poison	318	45	49	49	65
2	Ivysaur	Grass	Poison	405	60	62	63	80
3	Venusaur	Grass	Poison	525	80	82	83	100
3	VenusaurMega Venu	Grass	Poison	625	80	100	123	122
4	Charmander	Fire		309	39	52	43	60
5	Chameleon	Fire		405	58	64	58	80

Figura 2: Mesmo Pokémon (números da primeira coluna são iguais), mas com status diferentes.

O banco de dados não possui missing data, contudo há casos em que a coluna 'Type 2' aparece vazia (Figura 3), isso é explicado porque muitos Pokémon só possuem um tipo e o autor da tabela decidiu deixar o espaço do 'Type 2' em branco.

#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk
1	Bulbasaur	Grass	Poison	318	45	49	49	65
2	Ivysaur	Grass	Poison	405	60	62	63	80
3	Venusaur	Grass	Poison	525	80	82	83	100
3	VenusaurMega Venu	Grass	Poison	625	80	100	123	122
4	Charmander	Fire		309	39	52	43	60
5	Chameleon	Fire		405	58	64	58	80

Figura 3: Pokémon com um tipo apenas, deixando a coluna 'Type 2' vazia

De início temos 13 colunas, mas para nosso trabalho não foram utilizadas as colunas referentes ao id do Pokémon (primeira coluna) e ao seu nome (segunda coluna). Temos então 7 variáveis quantitativas discretas: Pontos de Vida, Ataque, Defesa, Ataque Especial, Defesa Especial, Velocidade e Geração. E 2 variáveis qualitativas nominais: Tipos 1 e 2, Lendários.

## 1.1 Descrição das variáveis:

Variáveis quantitativas discretas:

- Pontos de Vida (HP): define quanto de dano o Pokémon consegue suportar antes de ficar debilitado. Quanto maior, mais dano poderá levar.
- Ataque (Attack): modificador base para ataques físicos. Quanto maior, mais pontos de vida irá tirar do adversário.
- Defesa (Defense): resistência contra ataques físicos. Quanto maior, menor o número de pontos de vida que irá perder.
- Ataque Especial (Sp. Atk): modificador para ataques mágicos. Assim como Ataque, quanto maior, mais pontos de vida irá tirar do adversário.
- Defesa Especial (Sp. Def): modificador para a resistência a ataques mágicos. Quanto maior, menos pontos de vida irá perder.
- Velocidade (Speed): determina qual Pokémon ataca primeiro em cada turno, também determina sua chance de acertar um ataque ou de esquivar de um.
- Geração (Generation): a cada jogo lançado surge com ele uma nova leva de Pokémon e cada uma dessas leva é chamada de geração.

Variáveis qualitativas nominais:

- Tipos 1 e 2 (Type 1, Type2): cada Pokémon tem um tipo, sendo que, esse tipo determina suas fraquezas e vantagens.
- Lendário (Legendary): definido com o valor 'TRUE' ou 'FALSE' ele indica se aquele Pokémon é um Pokémon lendário, sendo estes mais fortes e raros do que Pokémon comuns.

## 2. Código em R:

Para gerar os gráficos foi usado a biblioteca 'ggplot2'. Alguns códigos usados para gerar os gráficos.

### **Código para gerar os *BoxPlots*:**

```
library(ggplot2)

nomes = colnames(Pokémon)

nome = ggplot(Pokémon, aes(factor(Generation), Total)) + geom_boxplot(aes(colour =
factor(Generation)))

nome = nome + xlab("Generation")+ylab("Total")+labs(colour = "Generation")

print(nome)

for (b in 6:11){
  y = Pokémon[b];
  nome = ggplot(Pokémon, aes(factor(Generation), y)) + geom_boxplot(aes(colour =
factor(Generation)))
  nome = nome + xlab("Generation") + ylab(nomes[b]) + labs(colour = "Generation")
  print(nome)
}
```

### **Código para gerar o gráfico do cruzamento dos tipos e a quantidades de Pokémon em cada um deles:**

```
library(ggplot2)

nomes = colnames(Type)

nome = ggplot(Pokémon, aes(factor(Type2))) + geom_(aes(fill =factor(Type1)))
nome=nome+xlabs("Types")+ylabs("")+labs(fill="Types")
theme(axis.title.x=element_blank(),axis.ticks.x=element_blank())
print(nome)
```

### **Código para gerar o gráfico de dispersão entre ataque e defesa:**

```
library(ggplot2)

nomes = colnames(Pokémon)

nome = ggplot(Pokémon, aes(Attack, Defense))+geom_jitter(aes(colour =factor(Type1)))
nome = nome + xlab("Attack")+ylab("Defense")+labs(colour = "")
print(nome)
```

### 3. Gráficos:

Nesta seção serão apresentados e discutidos os gráficos demonstrados durante a apresentação em sala de aula. São 18 gráficos que relacionam diferentes características dos Pokémon e é possível observar certos padrões de comportamento nessas relações.

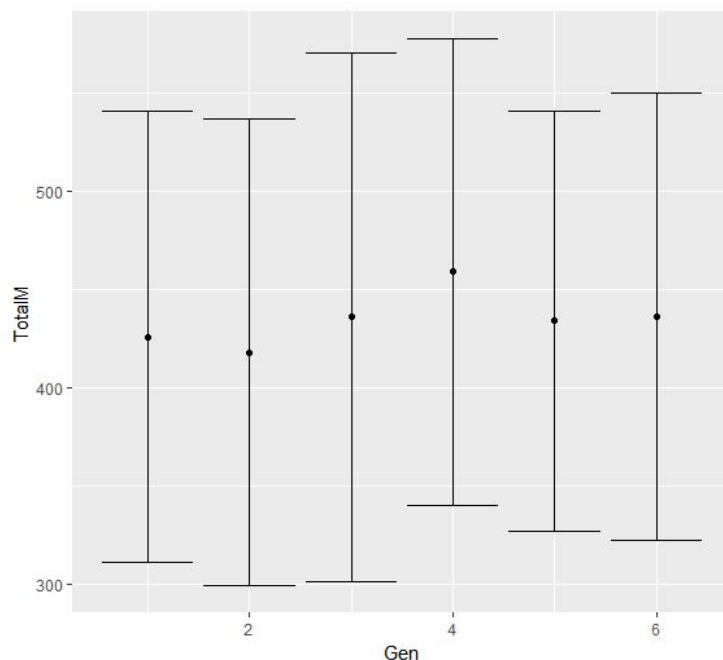


Gráfico 1: Média das somas de atributos por Geração.

Os gráficos 1 e 2 mostram que, de um modo geral, as gerações são equilibradas em termos médios de somas de atributos e pontos de vida. No gráfico 1, em especial, é possível observar alguns de exemplos de flutuações que se assemelham a padrões reguladores de atributos, ou seja, em gerações seguintes às que a soma dos atributos foi maior há uma queda da soma média, como se buscasse um equilíbrio entre as gerações. Pode-se supor que isso seja planejado pelos criadores do jogo com o intuito de balancear as gerações e tentar evitar preferências a determinadas gerações devido aos atributos.

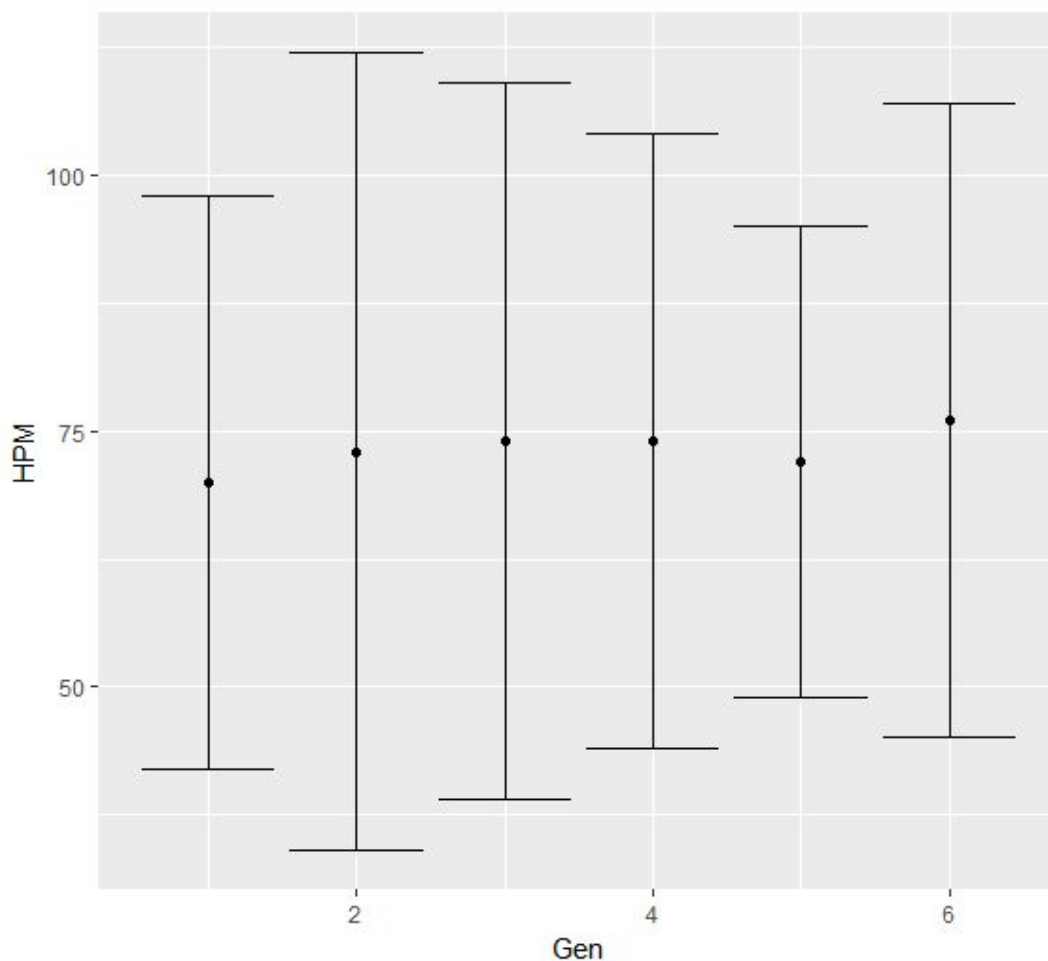


Gráfico 2: Média dos pontos de vida por Geração.

Outro aspecto interessante que pode ser notado é que a média da soma de atributos na quarta geração se sobressai levemente em relação às demais, mas isso não ocorre com as médias de pontos de vida, em que é a sexta geração que merece destaque, mesmo que por uma margem ínfima.

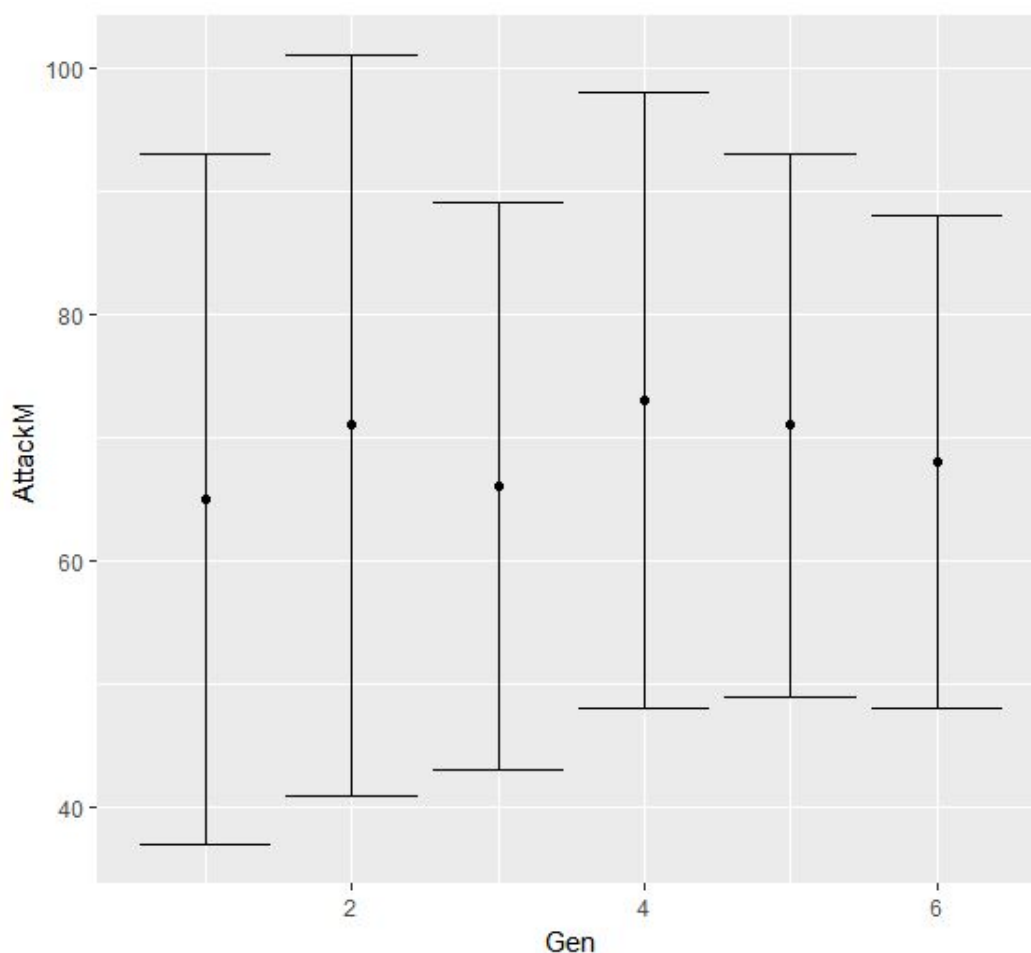


Gráfico 3: Médias de Ataque por Geração

Nos gráficos 3 e 4 observa-se um comportamento similar aos anteriores, no sentido de buscar um equilíbrio relativo entre as gerações. Porém, ainda que por pouco, as diferenças entre gerações são mais significativas, indicando que as modificações que poderiam estar buscando equilibrar o jogo precisam ser mais drásticas nos atributos Ataque e Defesa.

Além disso, comparando os dois gráficos, percebe-se que uma certa estabilidade entre as médias de Ataque e Defesa só foi obtida após a terceira geração, pois até então as diferenças entre as médias eram maiores.



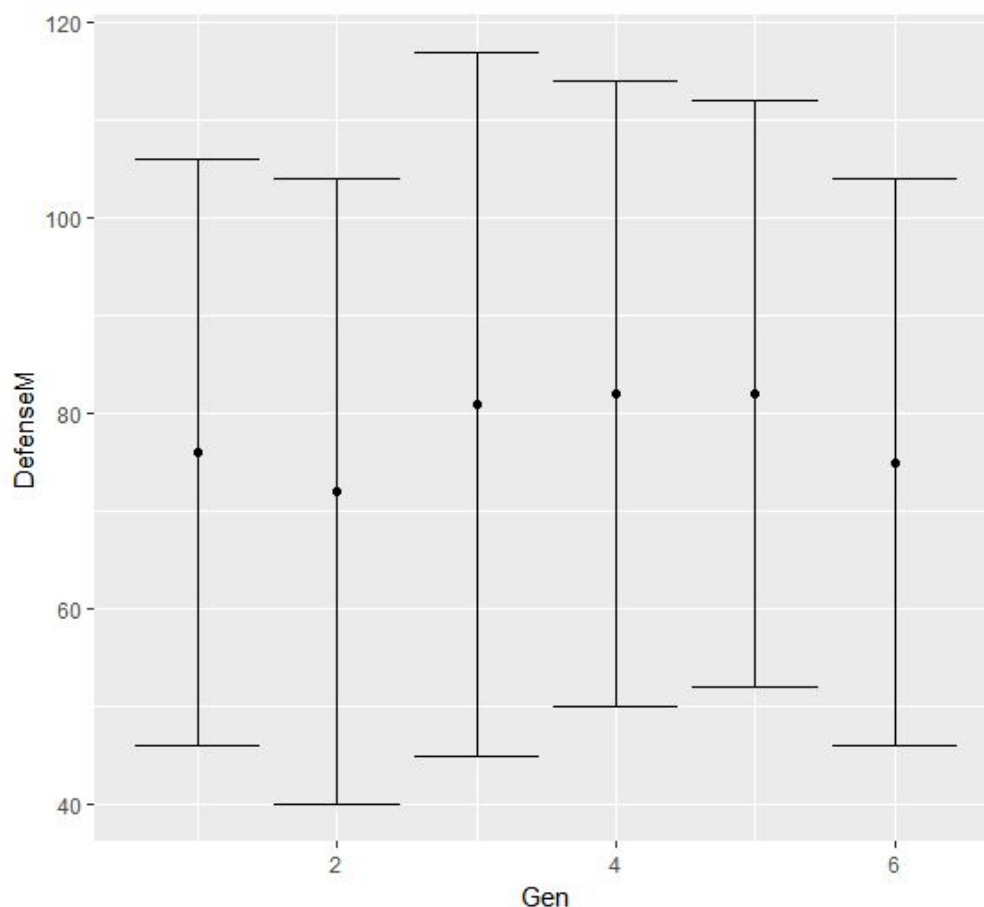


Gráfico 4: Médias de Defesa por Geração.

Os gráficos a seguir (5 à 8) são *boxplots* das variáveis anteriores. Nestes é possível analisar, visualmente, as medianas e desvios-padrão de cada um dos atributos em relação à geração e tentar encontrar mais padrões ou pontos específicos que fogem a estes (*outliers*).

No gráfico 5, nota-se que as medianas das somas totais de atributos por geração estão bastante próximas e que a imensa maioria dos dados concentra-se dentro da região do desvio-padrão, possuindo apenas um único *outlier*. Isto indica que os dados estão distribuídos de maneira bastante uniforme e que a média estará próxima da mediana em alguns casos.

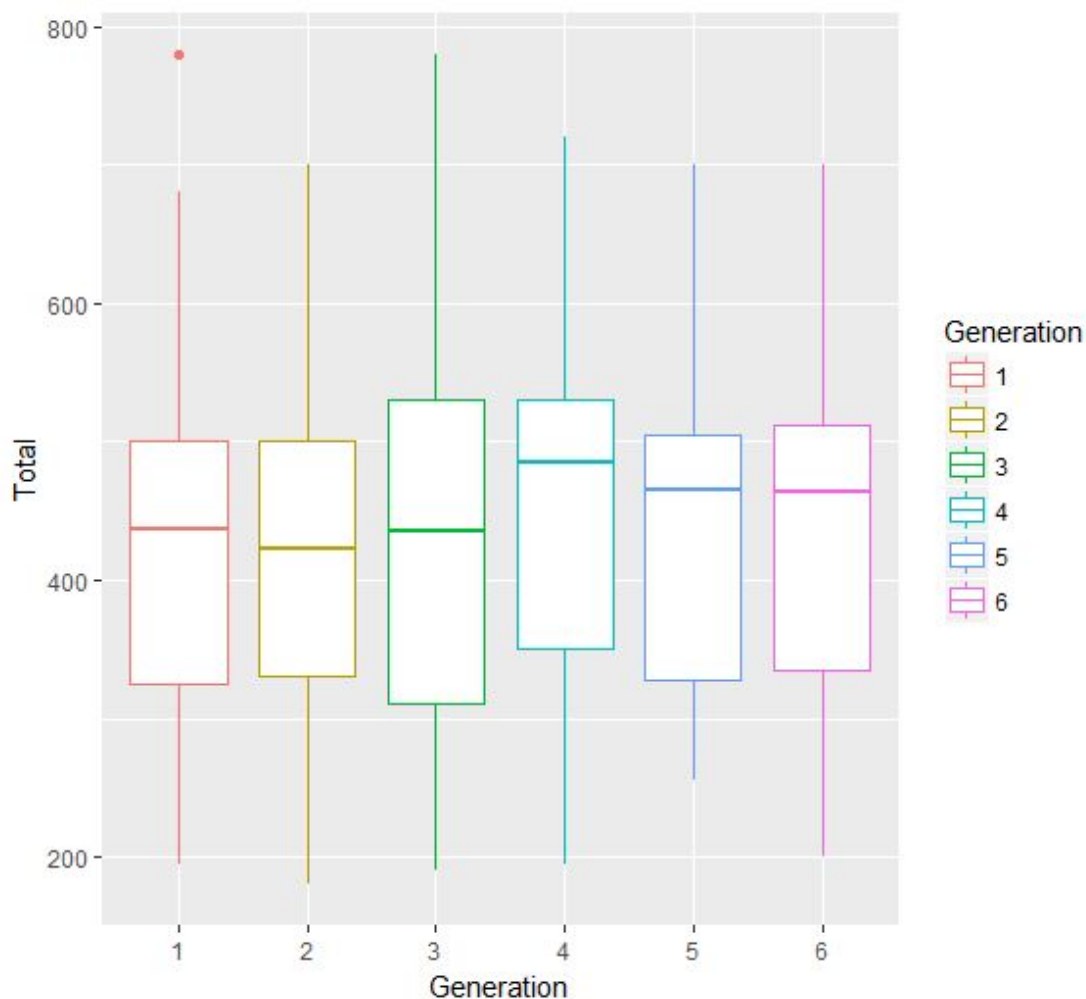


Gráfico 5: BoxPlot das Somas dos atributos por Geração

Mais uma vez é possível notar uma diferença comportamental entre as somas dos atributos e os pontos de vida. No gráfico 5, possuía apenas um *outlier*, já o gráfico 6, que também plota o atributo em função das gerações apresenta um aumento significativo na quantidade de *outliers*. Em relação ao total de observações, ainda é possível afirmar que esses dados estão bem distribuídos, apesar do comportamento discrepante em comparação ao outro atributo, soma de atributo. Inclusive, devido a valores menores de desvios-padrão, é possível afirmar que os valores, de uma maneira geral concentram-se próximos à mediana.

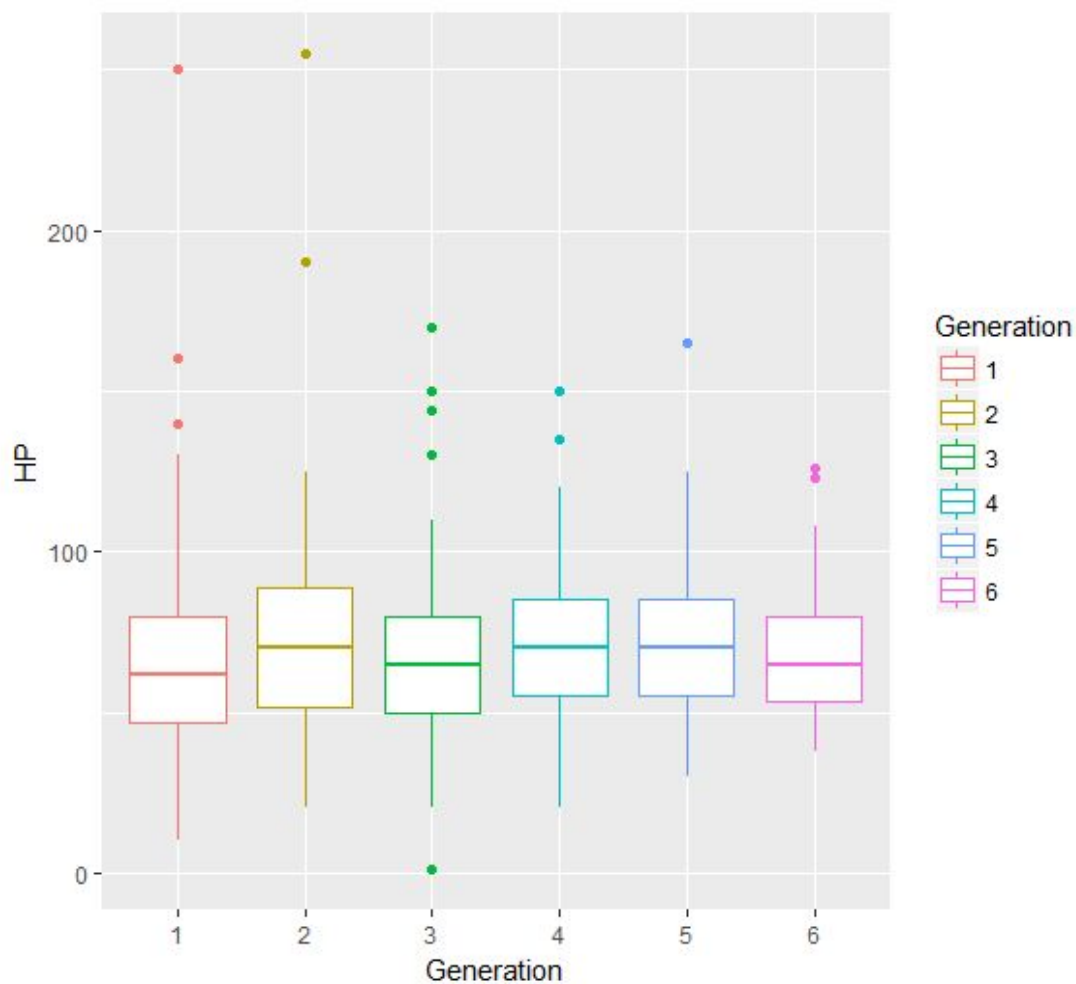


Gráfico 6: BoxPlot dos Pontos de Vida por Geração.

Outro ponto relevante é que as medianas entre as gerações estão muito próximas umas das outras, até mais próximas que as médias estavam no gráfico 4, indicando que provavelmente ao desconsiderar-se os *outliers*, o equilíbrio entre gerações seria ainda maior para este atributo.

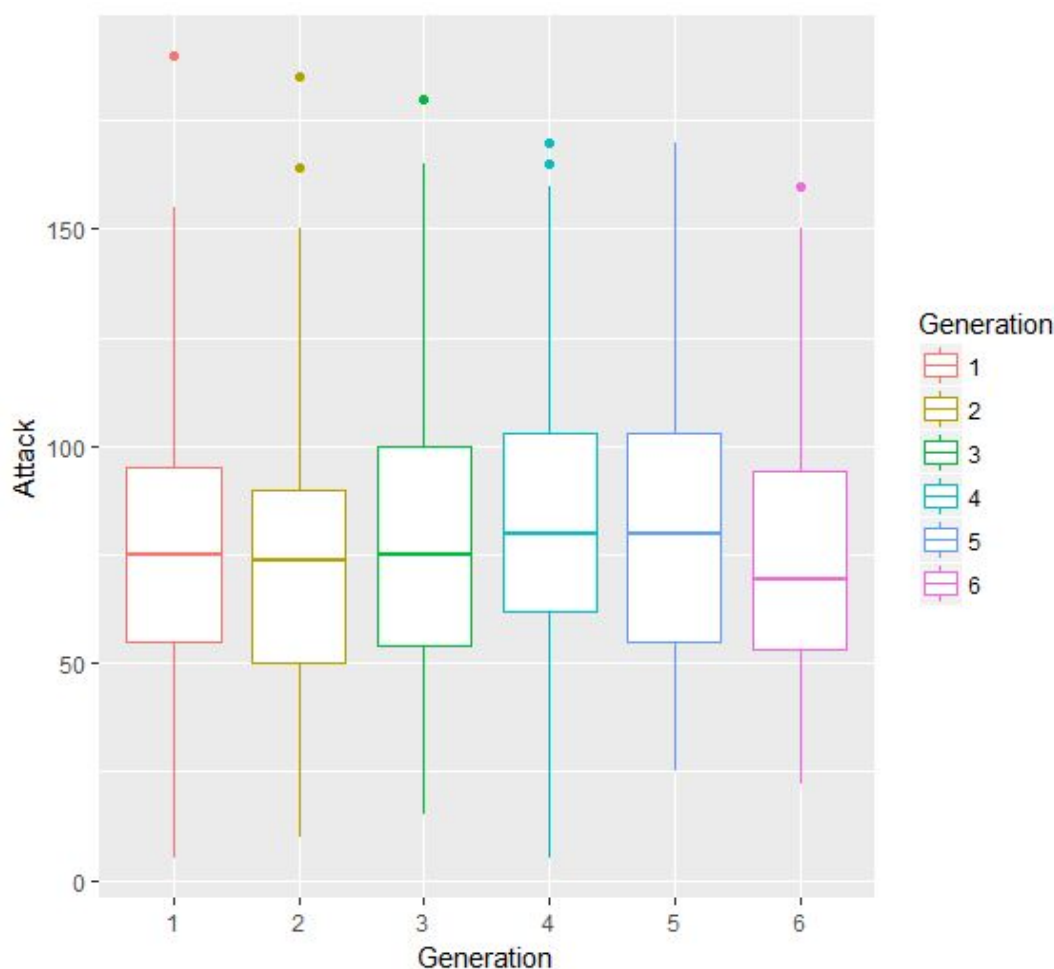


Gráfico 7: BoxPlot dos Ataques por Geração.

Diferente do gráfico 6, o número de *outliers* é mais reduzido no gráfico 7, mas os desvios-padrão são maiores e, portanto, a distribuição dos dados é um pouco mais dispersa, indicando que há uma gama maior de possibilidades para os valores de ataque do que para os pontos de vida, o que faz muito sentido no contexto do jogo em que o principal diferencial entre os Pokémon é sua habilidade de causar dano (Ataque ou Ataque Especial) ou de anular dano (Defesa ou Defesa Especial) e, em geral, não são os pontos de vida, salvo algumas exceções (nitidamente, os *outliers*).

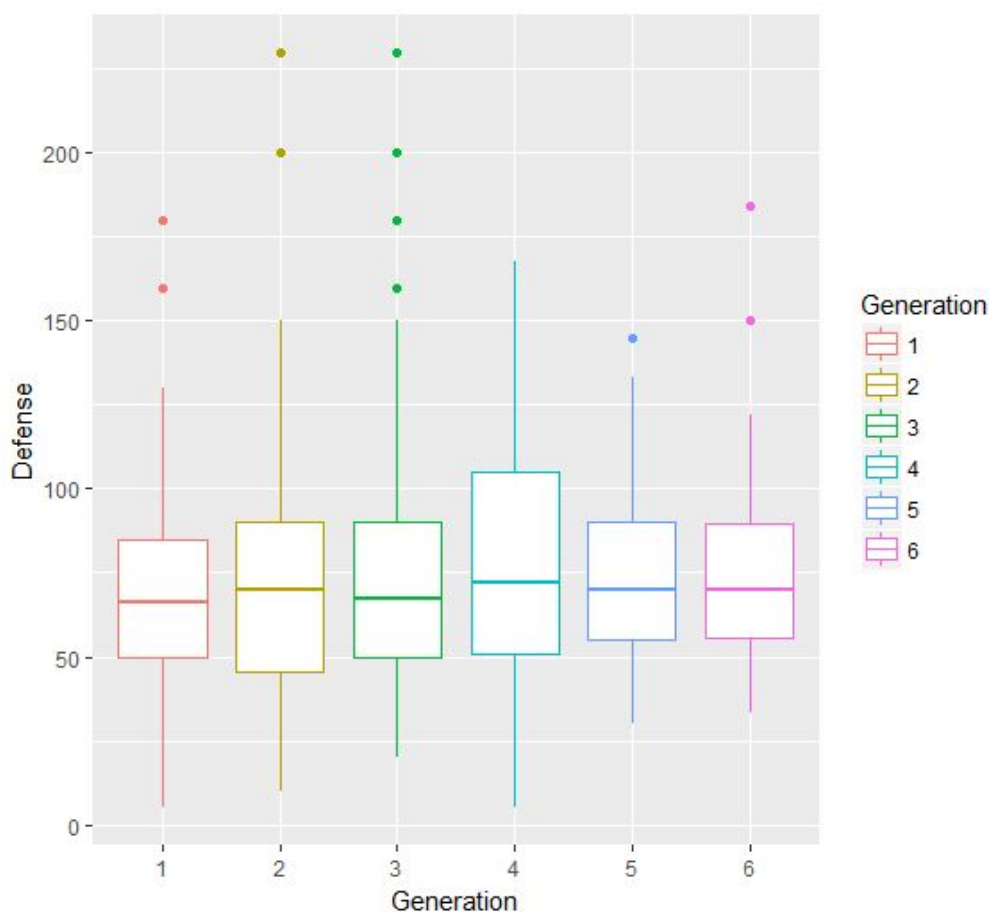


Gráfico 8: BoxPlot das Defesas por Geração.

Neste último *BoxPlot* (gráfico 8) é visível a semelhança entre as medianas e há poucos *outliers* em relação ao total de observações. Comparando os desvios-padrão deste com os anteriores, percebe-se que a variação é baixa e que apenas as flutuações de pontos de vida são menores. Isso indica que, potencialmente, os pontos de ataque são um pouco mais decisivos em combate que os pontos de defesa.

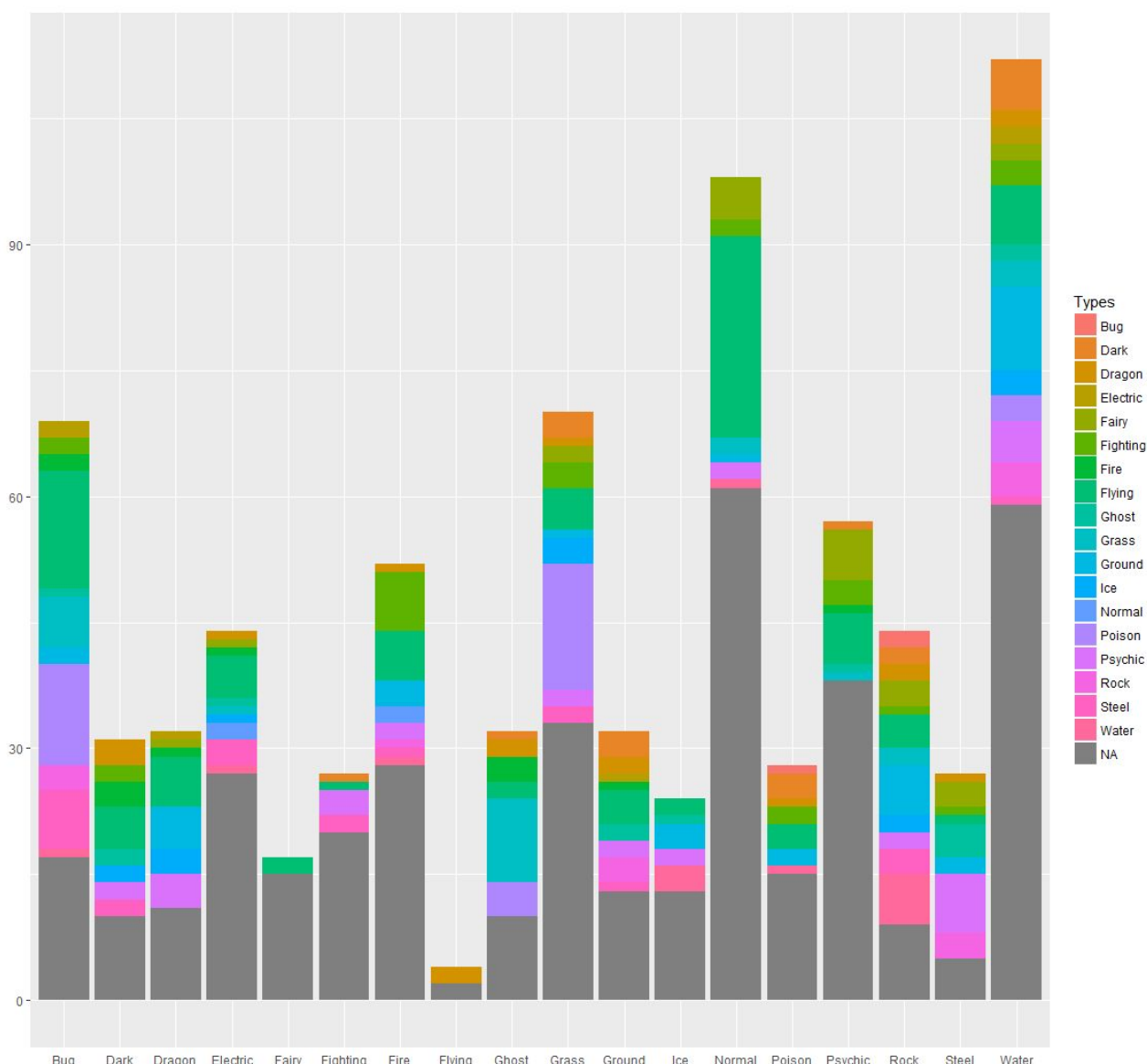


Gráfico 9: Representação em barras dos Tipos 1 e 2. No eixo x estão os Tipos 1 nos eixos y são as quantidades totais de Pokémon daquele tipo e as cores representam quantos possuem algum Tipo 2 associado ou não.

A primeira informação que o gráfico 9 remete é a de que a maioria dos Pokémon possui apenas o tipo 1. Como os tipos estão atrelados a características físicas dos Pokémon, não é possível, de imediato inferir algo sobre essa informação, mas ao longo de um processo analítico mais detalhado, nas próximas etapas do trabalho, será possível relacionar esta estatística com algum outro dado relevante.

O tipo 1 de maior ocorrência, para a surpresa do grupo, foi o tipo aquático (*water*). Esperava-se, por experiências pessoais com o jogo que tipo mais frequente fosse o normal, mas estatísticas mostraram algo diferente.

Outra informação interessante é que muito poucos Pokémon têm o tipo voador (*Flying*) como tipo 1, mas ele muitos possuem este como tipo 2. Possivelmente haverá uma relação entre afinidade de tipos que explique isso, mas com essas estatísticas não é possível analisar desta forma.

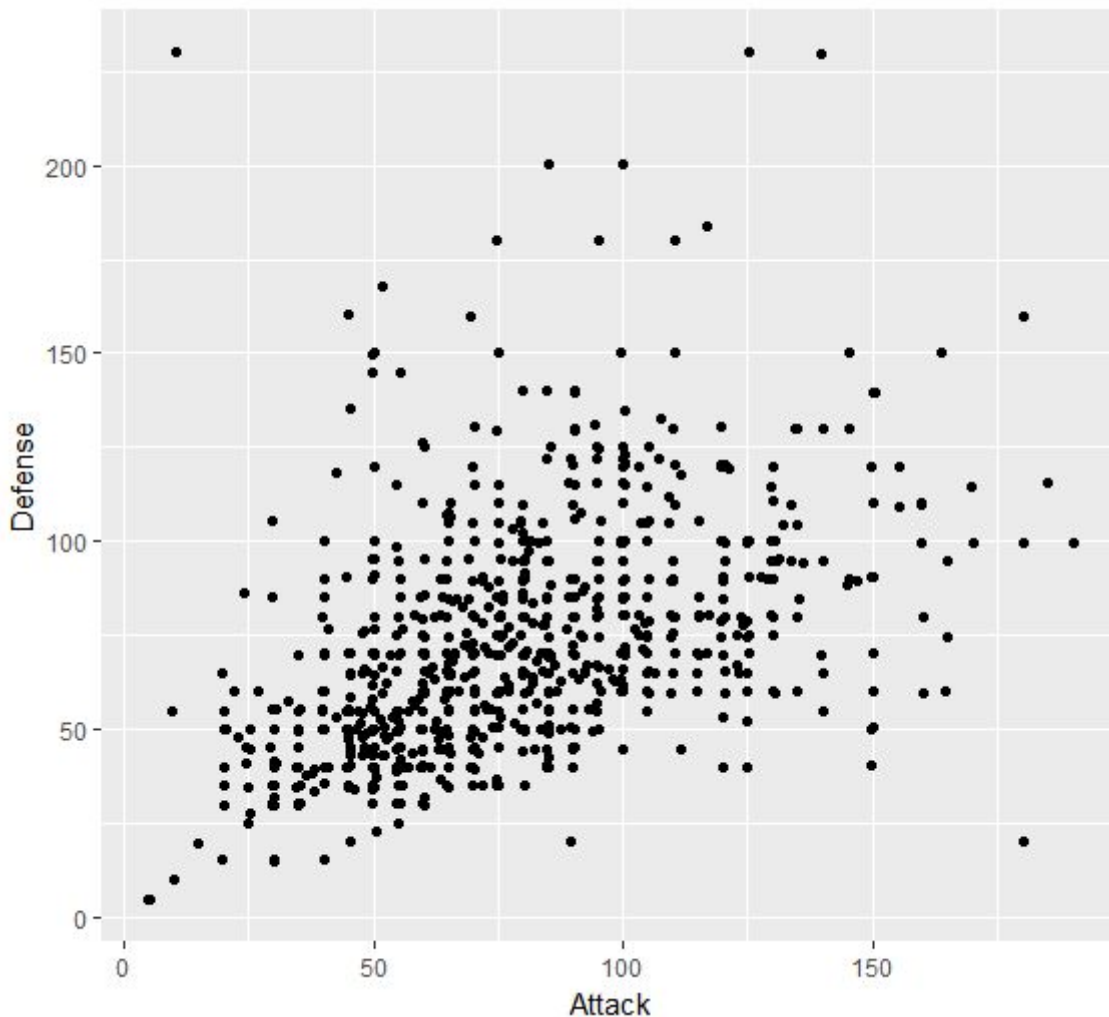


Gráfico 10: Ataques pelas Defesas de cada Pokémon.

No gráfico 10, é clara a dispersão dos dados, mas é possível perceber um padrão crescente entre ataque e defesa. De uma maneira geral, conforme cresce o ataque, cresce a defesa, mas não há Pokémon com defesa e ataque simultaneamente muito acima da média.

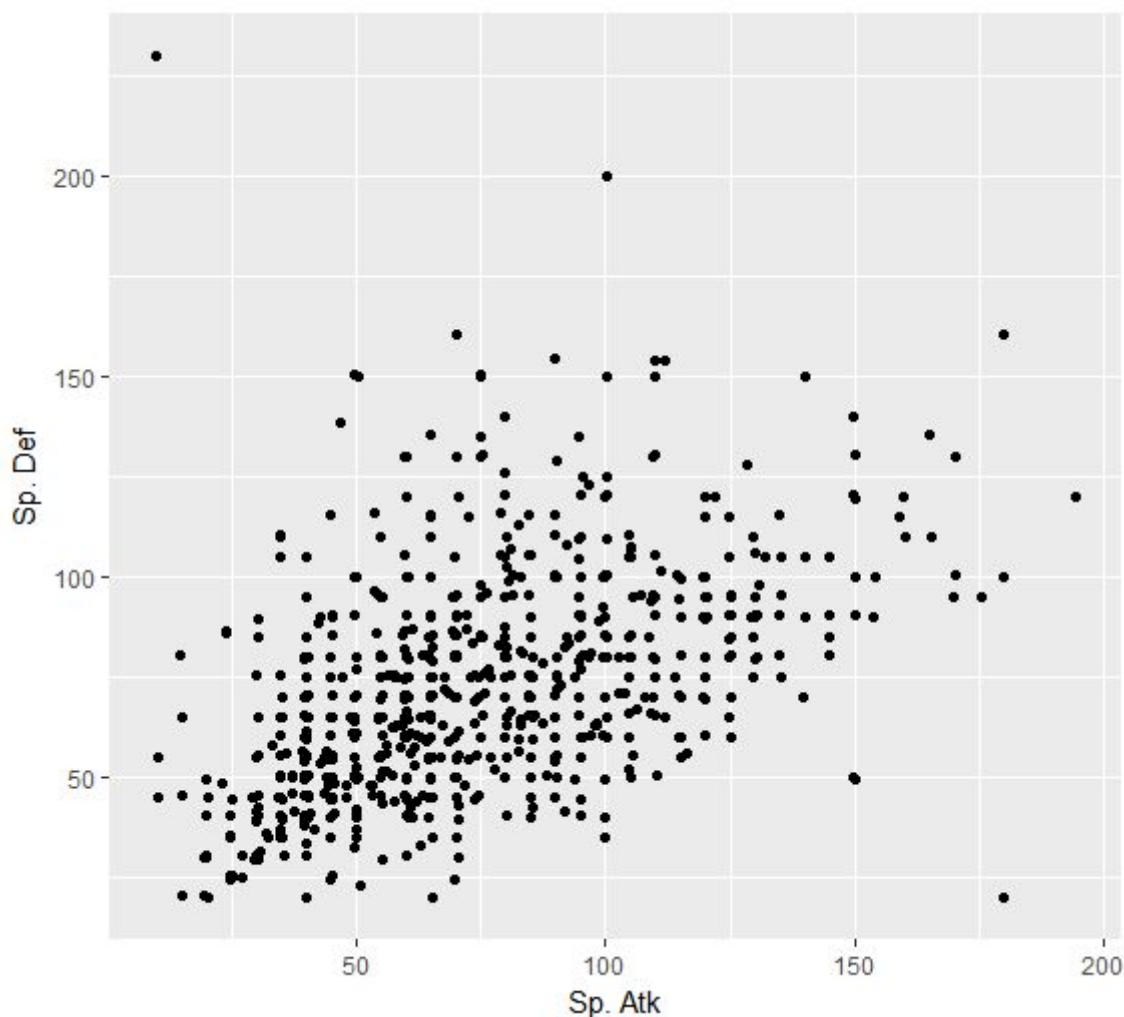
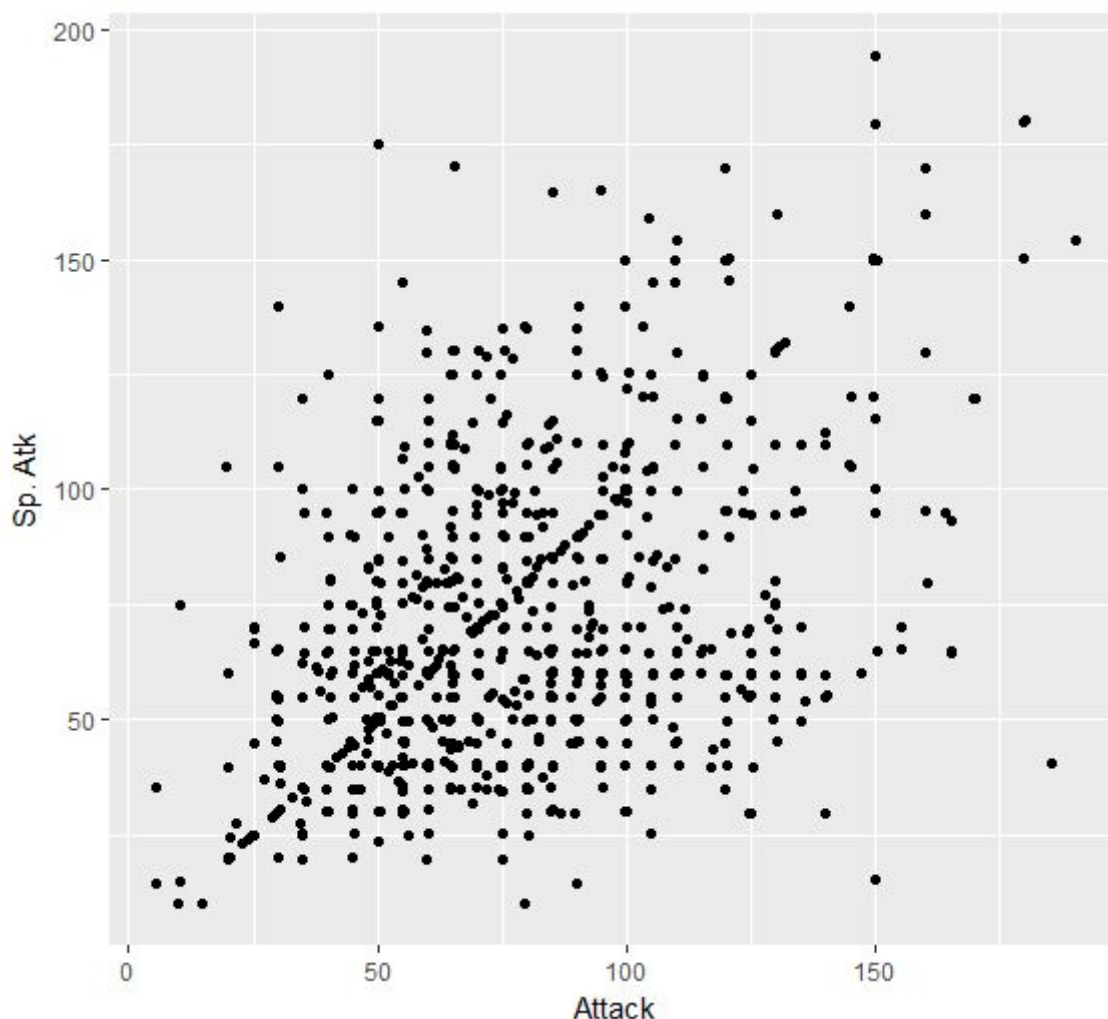


Gráfico 11: Ataques Especiais pelas Defesas Especiais para cada Pokémon.

Há um padrão similar ao gráfico 10 no 11, pois na relação de ataque especial e defesa especial também há uma certa dispersão dos dados e, de uma maneira geral, conforme cresce o ataque especial, cresce a defesa especial. Porém, neste caso, a defesa especial cresce em uma taxa menor em relação ao ataque especial do que cresce a defesa em relação ao ataque.





*Gráfico 12: Ataques pelos Ataques Especiais para cada Pokémon.*

Em ambos os gráficos 12 e 13 é possível notar claramente uma tendência linear no comportamento da função, pois há uma faixa de dados agrupada no centro dos gráficos. Este comportamento indica que a relações de ataque com ataque especial e defesa com defesa especial são bastante proporcionais e que a taxa de crescimento dos ataques está intrinsecamente relacionada à taxa de crescimento dos ataques especiais e vice-versa. O mesmo pode-se afirmar acerca dos atributos representados no gráfico 13.

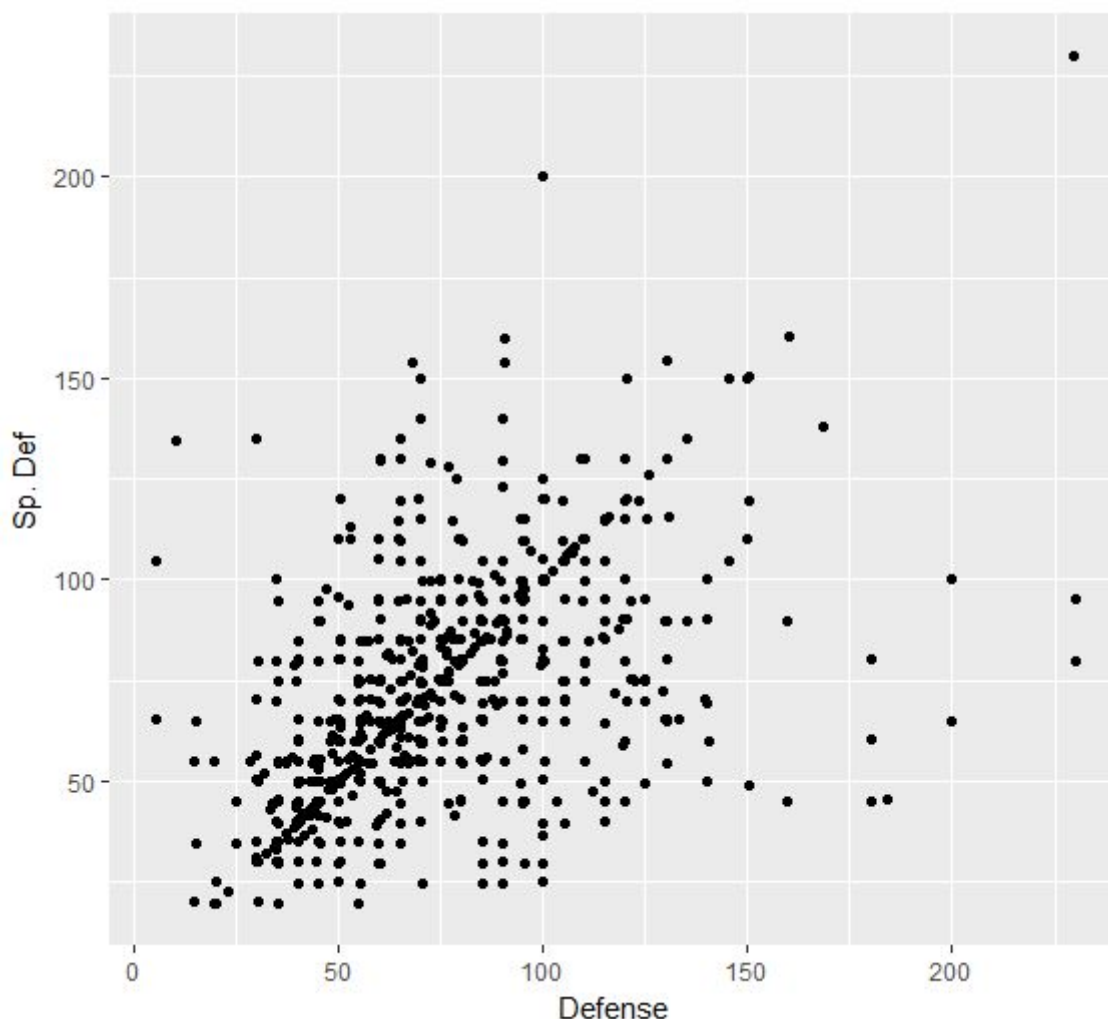
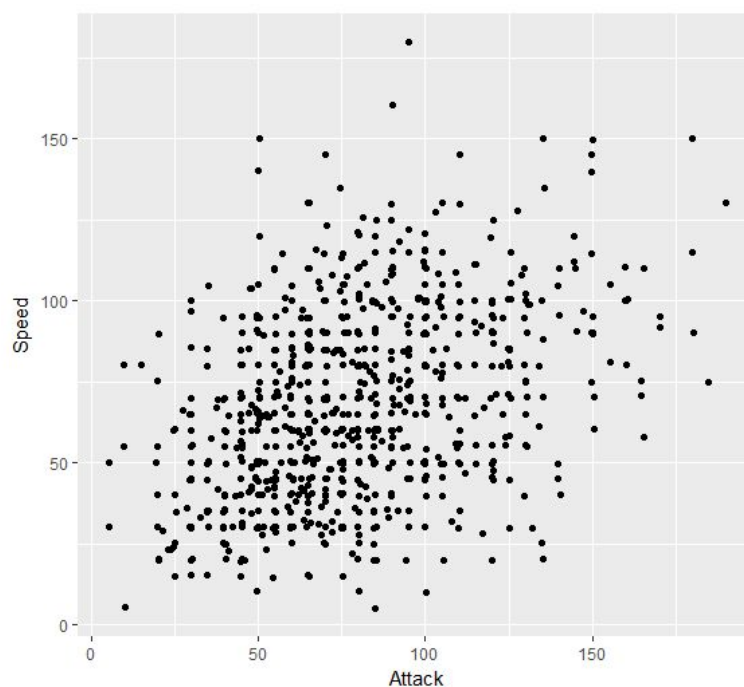
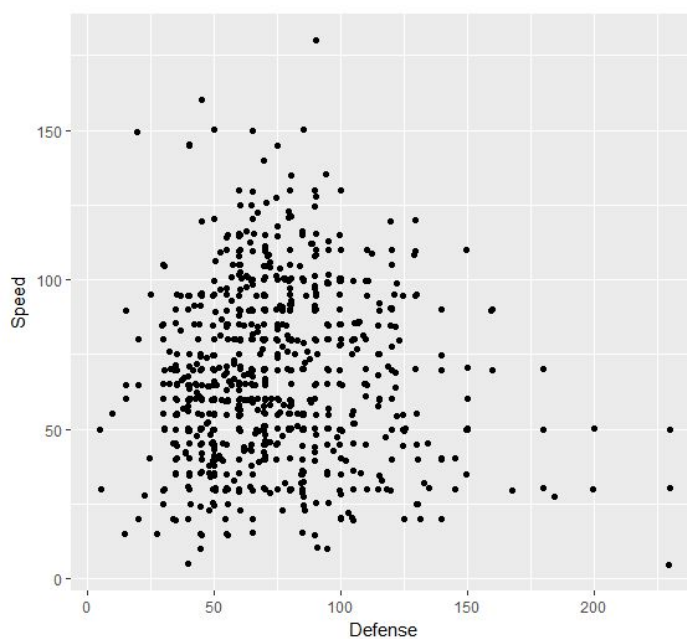


Gráfico 13: Defesas pelas Defesas Especiais para cada Pokémon.

Os gráficos 14 e 15 na página seguinte relacionam as variáveis de ataque e defesa com as velocidades. Pelo alto grau de dispersão visual dos dados foi possível inferir que há pouca correlação entre a velocidade e qualquer dos dois atributos, indicando que melhorias ou reduções na velocidade não implicam em respectivas mudanças nos valores dos atributos de ataque e defesa. Como o comportamento de ataques e defesas especiais já foi determinado como intrínseco aos ataques e defesas, pode-se afirmar que as conclusões acima valem também para estes atributos.



*Gráfico 14: Ataques pelas Velocidades para cada Pokémon.*



*Gráfico 15: Defesas pelas Velocidades para cada Pokémon.*

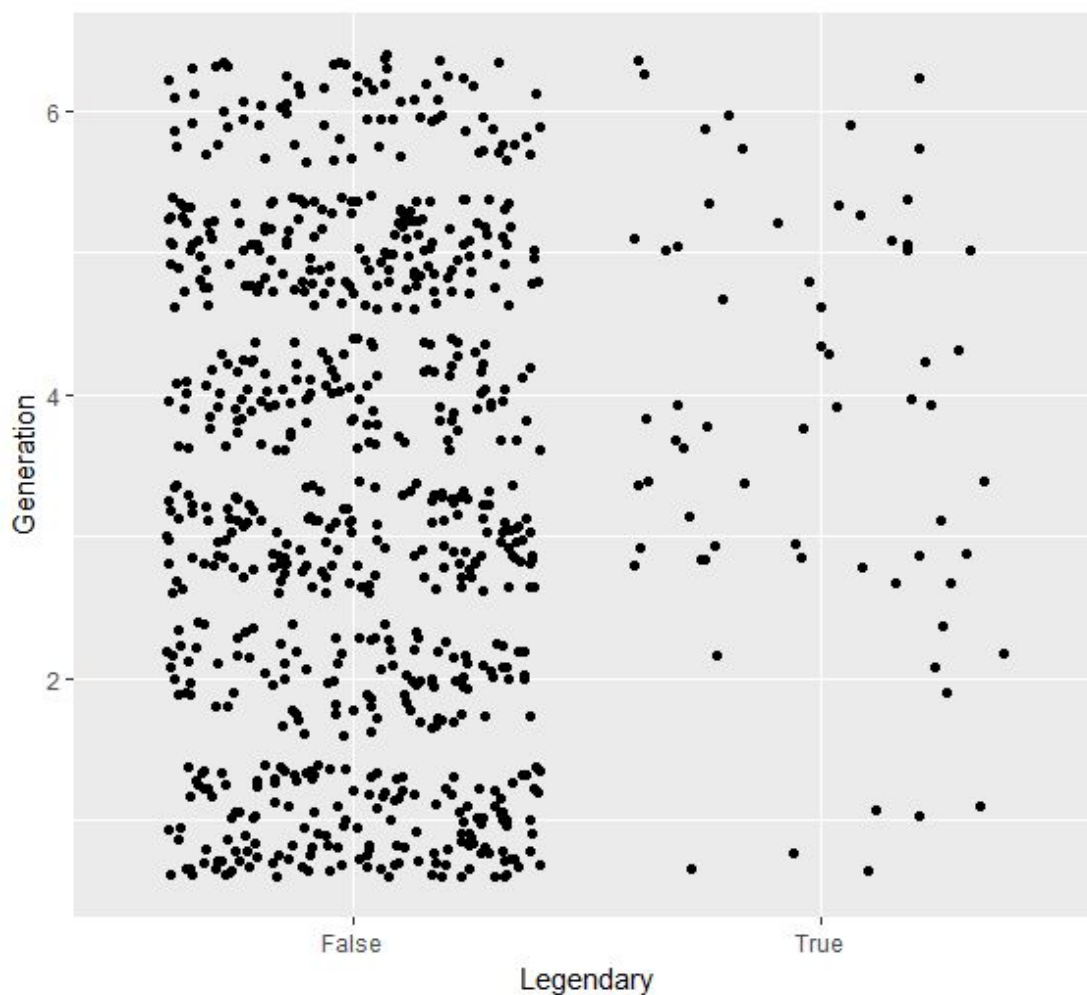


Gráfico 16: Lendários por Geração.

Neste gráfico (16) é perceptível a diferença entre o número de lendários por geração. Nas primeiras gerações (1 e 2) a quantidade total de Pokémon lendários era menor que nas gerações seguintes. Isto é um indicativo de que houve, por algum motivo, a necessidade de introduzir mais Pokémon com a característica lendário, ao longo das gerações. Especulando, podemos dizer que os criadores perceberam uma grande aceitação do público-alvo em relação a este tipo especial de Pokémon e buscaram trazer mais para que houvesse um atendimento melhor à demanda.

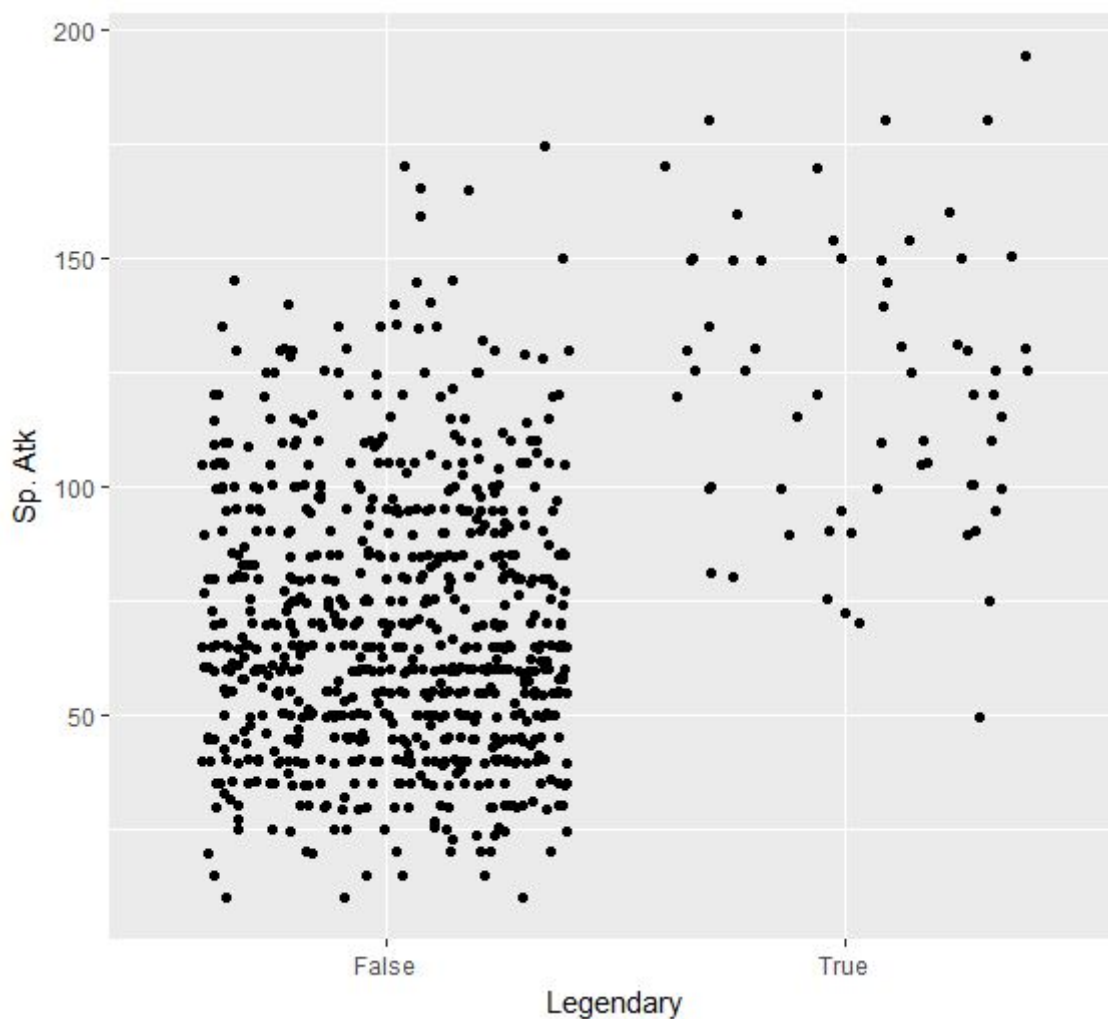


Gráfico 17: Lendários por Ataque Especial.

No gráfico 17, claramente os Pokémon lendários possuem uma vantagem significativa em ataques especiais, pois não só em média seus ataques são superiores, mas os maiores valores de ataques especiais estão no grupo dos lendários.

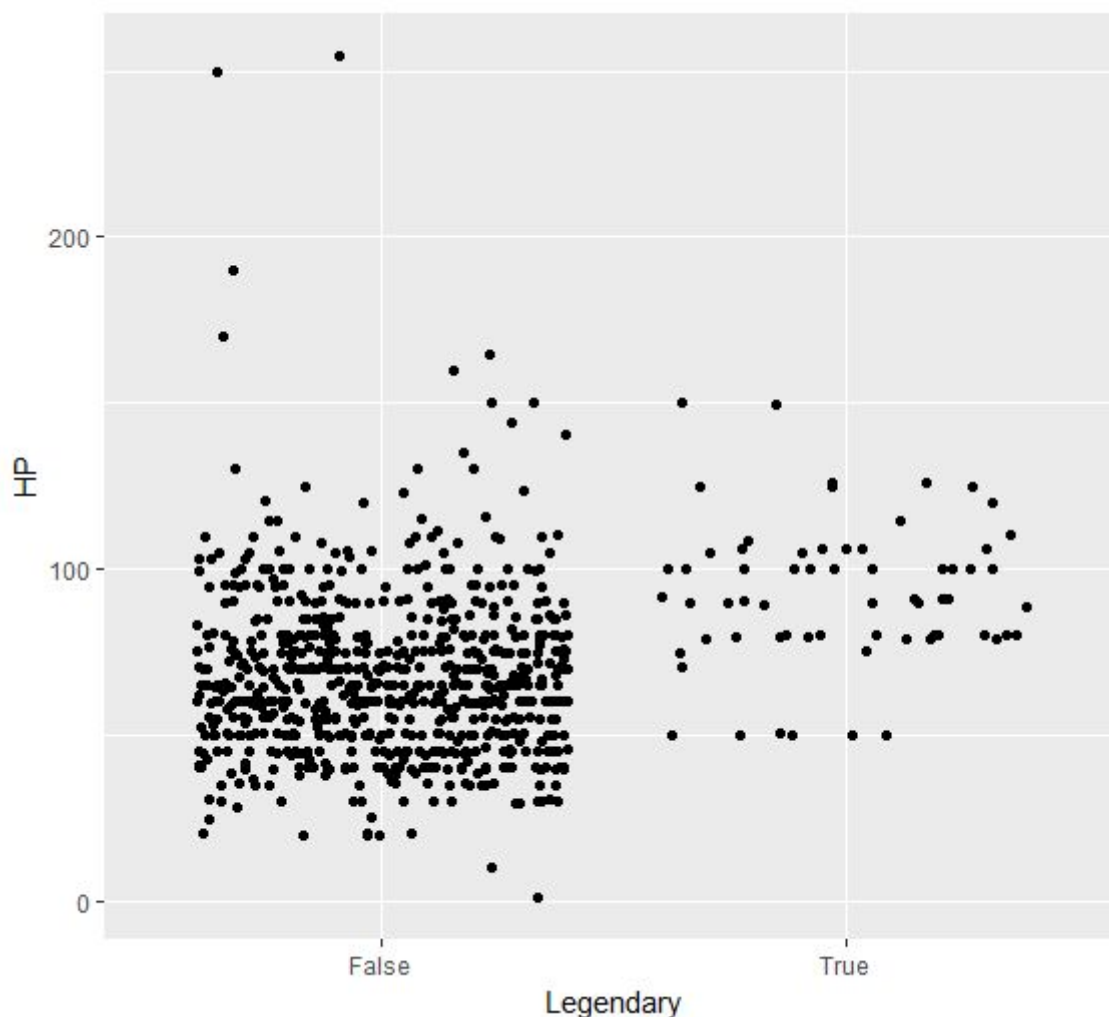


Gráfico 18: Lendários por Pontos de Vida.

Diferente do gráfico 17, o gráfico 18 possui um grau de dispersão muito inferior, pois nitidamente os dados estão concentrados, reforçando o que foi percebido no *BoxPlot* do gráfico 6, em que os desvio-padrão dos pontos de vida em cada geração eram baixos.

Do ponto de vista dos lendários, novamente, há uma vantagem em relação à média, porém os Pokémon com os maiores valores de pontos de vida não são do tipo lendário neste caso. Isto também reforça uma afirmação feita anteriormente de que os pontos de vida são menos significativos em batalha que outros atributos

#### **4. Próximos passos**

Ao analisarmos os gráficos gerados pelo banco de dados, concluimos que não há muitas diferenças entre as gerações, o que é mostrado pelas médias e box plots de variáveis dispostas por geração. Tendo notado esse fato, e conforme sugestão vista em sala de aula, nossos próximos esforços estarão focados em analisar as variáveis de acordo com os tipos de pokémon.

Com isso, pretendemos verificar se os atributos de um pokémon de determinado tipo diferem de um pokémon de outro tipo, e como esses atributos estão distribuídos de acordo com os tipos, atentando-nos para as relações estabelecidas entre pokémon de diferentes tipos. Também pretendemos observar as relações de cada variável de acordo com cada tipo de pokémon, e como elas são influenciadas por esses tipos.

Quanto às gerações, pretendemos utilizá-las como uma ordem temporal, com a finalidade de averiguar se as diferenças e semelhanças entre os tipos mantêm-se conforme as gerações passam.