



Instituto de Educação Superior de Brasília - IESB
Ciência de Dados e Inteligência Artificial

Análise exploratória sobre venda de carros na Noruega

por

Victor Augusto Souza Resende

Brasília - DF, 8 de Agosto de 2020

Vendas de Carros na Noruega

Trabalho para aplicação de técnicas estudadas no segundo semestre do curso de Ciência de Dados & Inteligência Artificial na matéria **Estatística**, juntamente com estudos particulares do autor sobre séries temporais e modelos simples de previsões ensinados no Curso de Inverno na matéria **R na prática**, ambos cursos citados ministrados pelo Instituto de Educação Superior de Brasília.

Dedicatória

À minha família, em especial aos meus pais,
minha namorada e meus amigos, pessoas da
qual obtive muito apoio.

Agradecimentos

Primeiramente, agradeço ao meu primo Marcelo e ao meu professor Suélio por tornar a documentação desse trabalho viável via LaTeX. Agradeço também à minha família e minha namorada Stefany pelo apoio e aos meus professores por grandes ensinamentos.

RESUMO

RESENDE, V.A. **Análise exploratória sobre venda de carros na Noruega nos anos de 2007 até janeiro de 2017.** 2020. Ciência de Dados Inteligência Artificial, Instituto de Educação Superior de Brasília, Brasília, 2020.

A Noruega por ser um país com grande facilidade em questões burocráticas tornou viável a coleta de dados apresentado nesse trabalho, e visando aplicar ensinamentos conquistados cursando até então o segundo semestre de Ciência de Dados Inteligência Artificial, foi efetuada uma análise exploratória sobre tal tema para encontrar pontos específicos e intuições de negócios com o uso de técnicas estatísticas como séries temporais e previsões com uso das ferramentas R e RStudio. Nesse trabalho foi tentado provar que a Noruega realmente está se tornando um país com carros majoritários elétricos analisando a venda de carros tradicionais. Toda a base de dados, intitulada como "New Car Sales Norway", se encontra disponível na plataforma Kaggle. Esse trabalho e o código utilizado se encontram no repositório do author na plataforma GitHub: github.com/victoresende19.

Palavras-chave: Análise Exploratória de Dados, R, RStudio, Venda de Carros, Noruega.

Abstract

RESENDE, V.A. **Exploratory analysis on car sales in Norway from 2007 to January of 2017**. 2020. Data Science Artificial Intelligence, Brasília Institute of Higher Education, Brasília, 2020.

Norway, being a country with great ease in bureaucratic issues, made the data collection presented in this work viable, and in order to apply lessons learned while studying the third semester of Data Science and Intelligence Artificial, an exploratory analysis was carried out on this topic to find specific points and business sights using statistical techniques such as time series and forecasts using the R and RStudio tools. In this work, the author tried to prove that Norway is really becoming a country with electric majority cars analyzing the sale of traditional cars. The entire database, entitled "New Car Sales Norway", is available on the Kaggle platform. This document and the scrip used were searched in the author's repository on the GitHub platform: github.com/victoresende19.

Key-words: Exploratory Analysis, R, RStudio, Car Sales, Norway.

Sumário

1	Conhecendo a Base de Dados	3
1.1	Introdução	3
1.2	Dicionário da Base de Dados	4
1.2.1	Notas sobre as variáveis	5
1.2.2	Notas sobre a base de dados	5
2	Limpeza dos dados	6
2.1	Quantidade de Dados faltantes	6
2.2	Excluindo Dados faltantes	7
3	Exploração de Dados - Estatísticas Descritivas	8
3.1	Medidas de tendência central	8
3.1.1	Medidas de tendência central - Quantidade	8
3.1.2	Medidas de tendência central - Quantidade & Marca	9
3.1.3	Medidas de tendência central - Importações	10
3.1.4	Box Plot	11
3.2	Medidas de dispersão	13
3.2.1	Medidas de dispersão - Quantidade	14
3.2.2	Medidas de dispersão - Quantidade & Marca	14
3.2.3	Medidas de dispersão - Importações	15
3.3	Formato da distribuição dos dados	16
3.3.1	Formato da distribuição dos dados - Quantidade	17
3.3.2	Formato da distribuição dos dados - Quantidade & Marca	18
3.3.3	Formato da distribuição dos dados - Importações	20
3.4	Correlação entre as variáveis quantitativas	21
3.4.1	Ideia geral - Exemplo: Quantidade & Quantidade Diesel	21
3.4.2	Correlação das variáveis - Geral	23
4	Análise das medidas encontradas	26
4.1	Introdução às análises	26
4.1.1	Análise - Quantidade	26
4.1.2	Análise - Quantidade & Marca	27
4.1.3	Análise - Importação	28
5	Regressões e Previsões	29
5.1	Regressão Linear	29
5.2	Regressão Linear - CO2 vendidos vs. Ano	30
5.3	Regressão Linear - Quantidade vendidos vs. Ano	31

6	Série Temporais - Análise	33
6.1	Séries Temporais	33
6.2	Série Temporal - Quantidade vs. Ano	33
6.2.1	Série Temporal - Quantidade vs. Mês e Ano	34
7	Previsão - Análises	35
7.1	Previsão	35
7.1.1	Previsão com Holt-Winters	36
7.1.2	Previsão com ARIMA	37
7.1.3	Previsão com NAIVE	38
7.2	Análise acurácias	39
8	Conclusão	41

Capítulo 1

Conhecendo a Base de Dados

1.1 Introdução

O setor automotivo de um país é seguramente um dos setores econômicos que certamente pode-se fazer diversas análises a serem refletidas na população. Por exemplo, se um país vende diversos veículos, isso pode dizer que, de certa maneira, a economia do país está crescendo frequentemente pois a população possui renda para efetuar a compra de tais veículos. Porém por outro lado, se um país tem queda no setor automotivo, podemos ter algumas alternativas nebulosas para se entender. Por exemplo, a população pode estar empobrecida e não possui dinheiro para efetuar a compra de veículos automotivos, o país pode estar enfrentando alguma crise econômica (Seja interna ou global), ou então, a população está escolhendo um caminho mais saudável em relação à natureza, como por exemplo, os cidadãos podem estar optando por bicicletas, se locomover por meio de transporte público ou alguma outra alternativa que seja menos degradante ao ambiente.

Nesse trabalho será analisado, por meio de métodos estatísticos e com auxílio de ferramentas de software, a venda de carros na Noruega e importação de veículos efetuado pelo país no período de ano 2007 até janeiro de 2017. Será analisado também especificamente algumas marcas de carros escolhidas aleatoriamente pelo autor usando artifícios como, por exemplo, visualização de correlação entre variáveis, regressões Lineares, séries temporais e modelos básicos de previsões. Nesse trabalho, procurou-se provar que a Noruega está se tornando um país com carros majoritariamente elétricos, analisando por meio de regressões lineares, séries temporais a venda de carros tradicionais (Emitentes de CO₂). Vale ressaltar que todas as aplicações passaram por estudo prévio e, posteriormente, implementação usando as ferramentas R e RStudio. É destacável que esse trabalho possui referências obtidas na leitura dos livros: Estatística Básica por Morettin e BUSSAB (2017), The R book por Crawley (2012), Data science for business por Provost e Fawcett (2013) e em reportagens abordadas pela revista EPOCA.

Este estudo foi concluído em apenas um mês apenas com o objetivo de aplicação de técnicas aprendidas, sendo assim o tempo de confecção foi levado muito em consideração na escolha das variáveis para que as análises pudessem ser o mais eficaz possível.

1.2 Dicionário da Base de Dados

A base de dados está disponível em:

Kaggle - New Cars Sales Norway.

Como todas as variáveis estão em inglês, aqui será explicado o significado de cada

- **Year:** Ano de venda.
- **Month:** Mês de venda.
- **Quantity:** Quantidade vendida de automóveis (Geral).
- **Quantity YoY:** Quantidade vendida de automóveis Ano a Ano.
- **Import:** Quantidade de importações de automóveis (Automóveis usados).
- **Import YoY:** Quantidade de importações de automóveis Ano a Ano.
- **Used:** Quantidade vendida de automóveis usados.
- **Used YoY:** Quantidade vendida de automóveis usados Ano a Ano.
- **Avg CO2:** Emissão média de CO2 de automóveis vendidos.
- **Bensin CO2:** Emissão média de CO2 de abastecimento com Bensina.
- **Diesel CO2:** Emissão média de CO2 de abastecimento com Diesel.
- **Quantity Diesel:** Quantidade de carros abastecidos com Diesel vendidos.
- **Diesel Share:** Quantidade de automóveis a Diesel vendidos.
- **Diesel Share Ly:** Quantidade de automóveis a Diesel vendidos há um ano.
- **Quantity Hybrid:** Quantidade de automóveis híbridos novos vendidos.
- **Quantity Eletric:** Quantidade de automóveis elétricos novos vendidos.
- **Import Eletric:** Quantidade de importações de automóveis elétricos usados.
- **Make:** Marca do automóvel vendido.
- **Model:** Modelo do automóvel vendido.

1.2.1 Notas sobre as variáveis

As seguintes variáveis: *Quantity Hybrid* e *Quantity Eletric* apenas começaram a serem postados a partir do ano 2011.

Em relação à variável *Diesel Share* foi utilizada a partir da seguinte fórmula: $\frac{Quantity_{Diesel}}{Quantity}$.

Como pôde ser evidenciado, a base de dados possui algumas variáveis com determinada denominação *YoY*. YoY é uma abreviatura para *Year Over Year*, ao pé da letra poderia ser traduzida para a língua portuguesa como Ano a Ano. YoY é uma medida usada para comparar crescimento de determinado seguimento atual em comparação ao mesmo seguimento num período passado, podendo assim calcular a taxa de crescimento ou decrescimento do seguimento analisado. Vale lembrar que tal medida é comumente retornada em porcentagem. Sendo assim temos que a formula de YoY é dada por:

$$\frac{PerodoAtual - PerodoPassado}{PerodoPassado} * 100.$$

É importante citar que nesse trabalho será avaliado de maneira mais específica as seguintes variáveis, pois este trabalho tem como objetivo apenas provar estudos realizados:

- Quantity
- Quantity & Make
- Import

Tal escolha se deu por uma melhor associação das variáveis com as análises que serão feitas juntamente com a ideia de serem efetuadas análises rasas porém eficientes, já que demandaria muito tempo efetuar as mesmas com a vasta gama de variáveis presentes na base de dados disponibilizada.

1.2.2 Notas sobre a base de dados

A base de dados é apresentada com todas as variáveis citadas anteriormente em relação à cada mês de acordo com determinado ano, como será apresentado na seguinte imagem:

	Year	Month	Quantity	Quantity_YoY	Import	Import_YoY	Used	Used_YoY	Avg_CO2	Bensin_Co2
1	2007	1	12685	5227	2276	257	NA	NA	152	155
2	2007	2	9793	2448	1992	-89	NA	NA	156	159
3	2007	3	11264	1445	2626	45	NA	NA	159	161
4	2007	4	8854	504	2220	-130	NA	NA	160	165
5	2007	5	12007	1592	2881	7	NA	NA	160	163
6	2007	6	11083	1545	3038	23	NA	NA	161	163
7	2007	7	12062	1908	3768	137	NA	NA	159	161
8	2007	8	10786	1993	3419	260	NA	NA	160	160
9	2007	9	9340	498	2897	-28	NA	NA	160	160
10	2007	10	11646	2973	3185	597	NA	NA	159	160
11	2007	11	10453	1709	2957	544	NA	NA	160	161
12	2007	12	9222	-1811	2097	-1257	NA	NA	162	162
13	2008	1	9901	-2784	2287	11	NA	NA	158	155
14	2008	2	10567	774	2627	635	NA	NA	160	159
15	2008	3	9506	-1758	2270	-356	NA	NA	159	160
16	2008	4	11704	2850	2930	710	NA	NA	159	160

Figura 1.1: Exemplo base de dados

Capítulo 2

Limpeza dos dados

2.1 Quantidade de Dados faltantes

Após conhecer as variáveis que serão trabalhadas, será efetuada a seguinte função na ferramenta RStudio para descobrir a porcentagem de dados faltantes em cada variável:

```
NAS <- round(colSums(is.na(norway_car_month))
              *100/nrow(norway_car_month),2)
```

Sendo assim, apenas foi necessário chamar a nova variável (*NAS*) da seguinte forma: *NAS > 0* para descobrir quais variáveis possuíam dados faltantes maior que zero. Portanto, temos o seguinte resultado:

- **Used:** 49.59% de dados faltantes.
- **Used YoY:** 59.50% de dados faltantes.
- **Quantity Hybrid:** 39.67% de dados faltantes.
- **Quantity Eletric:** 39.67% de dados faltantes.
- **Import Eletric:** 56.20% de dados faltantes.

2.2 Excluindo Dados faltantes

A ideia deste trabalho se da, principalmente, em tentar provar que a venda de carros elétricos aumentou durante os anos em relação à venda de carros tradicionais. Sendo assim, temos anteriormente que, por os carros elétricos serem inseridos apenas em 2011, existem muitos dados faltantes (em relação à todo período que essa base de dados contem), então foi analisado apenas os carros tradicionais e então refletir sobre a venda de carros elétricos/híbridos. Portanto assim, foram excluídas as seguintes variáveis:

- Quantity Hybrid
- Quantity Eletric
- Import Eletric

Sobrando assim apenas as seguintes variáveis:

- Used
- Used YoY

Essa abordagem foi decidida principalmente porquê seria arriscado usar técnicas de preenchimento podendo mudar o resultado final da análise.

A exclusão de cada variável citada anteriormente se deu com o seguinte código em R:

```
norway_car_month$Used_YoY <- NULL
norway_car_month$Used <- NULL
norway_car_month$Import_Electric <- NULL
norway_car_month$Quantity_Hybrid <- NULL
norway_car_month$Quantity_Electric <- NULL
```

Mesmo a Noruega, na data de confecção deste trabalho, tornando-se um país que vende mais carros elétricos do que tradicionais (De acordo com a líder da Associação Norueguesa de Veículos Elétricos, Christina Bu como foi relatado na revista EPOCA) ¹, não foi possível fazer tal análise de maneira eficiente pois a base de dados era referente até janeiro de 2017, no qual foi evidenciado que ainda possuía diversos dados faltantes, e o foco deste trabalho não são variáveis para carros híbridos/elétricos. Então, após analisar caso a caso visando o resultado final do trabalho, todas as variáveis que possuíam dados faltantes foram excluídas sem danos significativos à ideia final.

¹epocanegocios.globo.com/Mundo/noticia/2019/04/noruega-e-1-pais-do-mundo-vender-mais-carros-eletricos-do-que-veiculos-tradicionais.html

Capítulo 3

Exploração de Dados - Estatísticas Descritivas

3.1 Medidas de tendência central

As medidas de tendência central são medidas que representam o centro ou o meio de uma base de dados, algumas dessas medidas são: Média, Moda e Mediana. Porém nesse trabalho não iremos abordar de forma frequente a análise sobre a Moda pelo seguinte motivo:

- Para a variável Quantidade não faz sentido analisar a moda pois todo mês, durante os anos, teremos, certamente números de quantidade de vendas diferentes.
- Para a variável Quantidade e Modelo não faz sentido analisar a moda pois todo mês, durante os anos, teremos certamente números de quantidade de vendas diferentes.
- Para a variável Importação não faz sentido analisar a moda pois todo mês, durante os anos, teremos certamente números de importações diferentes.

3.1.1 Medidas de tendência central - Quantidade

A seguir será demonstrado o resultado das medidas de tendência centrais da variável *Quantidade* que se refere à quantidade de veículos vendidos nos período dos anos de 2007 até janeiro de 2017:

- **Média:** A média da quantidade de veículos vendidos em todos os meses nos anos de 2007 até janeiro de 2017 é aproximadamente 11134 veículos.
- **Mediana:** A mediana da quantidade de veículos vendidos em todos os meses nos anos de 2007 até janeiro de 2017 é de 11385 veículos.

Medidas alcançadas facilmente por meio do seguinte código na linguagem R usando a ferramenta RStudio:

```
mean(norway_car_month$Quantity)
11134.3
median(norway_car_month$Quantity)
11385
```

3.1.2 Medidas de tendência central - Quantidade & Marca

A seguir será demonstrado o resultado das medidas de tendência centrais da variável *Quantidade* em relação à variável *Marca* refletindo a quantidade de veículos vendidos em relação às marcas de veículos *Ford* e *Toyota*.

As seguintes medidas de tendência central para as variáveis *Quantidade* e *Ford* são:

- **Média:** A média de quantidade de veículos da **Ford** vendidos em todos os meses nos anos de 2007 até janeiro de 2017 é de 234 veículos.
- **Mediana:** A mediana de quantidade de veículos da **Ford** vendidos em todos os meses nos anos de 2007 até janeiro de 2017 é de 212 veículos.

As seguintes medidas de tendência central para a variável *Quantidade* e *Toyota* são:

- **Média:** A média de quantidade de veículos da **Toyota** vendidos em todos os meses nos anos de 2007 até janeiro de 2017 é de 273 veículos.
- **Mediana:** A mediana de quantidade de veículos da **Toyota** vendidos em todos os meses nos anos de 2007 até janeiro de 2017 é de 268 veículos.

Medidas alcançadas facilmente por meio do seguinte código na linguagem R usando a ferramenta RStudio:

```
#FORD
```

```
mean(norway_car_model$Quantity[norway_car_model$Make == 'Ford'])  
234.8089
```

```
median(norway_car_model$Quantity[norway_car_model$Make == 'Ford'])  
212
```

```
#TOYOTA
```

```
mean(norway_car_model$Quantity[norway_car_model$Make == 'Toyota'])  
273.6911
```

```
median(norway_car_model$Quantity[norway_car_model$Make == 'Toyota'])  
268
```


3.1.3 Medidas de tendência central - Importações

A seguir será demonstrado o resultado das medidas de tendência centrais da variável ***Importação*** que se refere à quantidade de veículos importados usado no período do ano de 2007 até janeiro de 2017.

- **Média:** A média da quantidade de veículos importados em todos os meses nos anos de 2007 até janeiro de 2017 é aproximadamente 2204 veículos.
- **Mediana:** A mediana da quantidade de veículos importados em todos os meses nos anos de 2007 até janeiro de 2017 é de 2263 veículos.

Medidas alcançadas facilmente por meio do seguinte código na linguagem R usando a ferramenta RStudio:

```
mean(norway_car_month$Import)  
2204.372
```

```
median(norway_car_month$Import)  
2263
```

3.1.4 Box Plot

A ferramenta estatística BoxPlot, presente também no software R e RStudio, pode nos trazer uma breve ilustração sobre medidas de tendência centrais, sendo representado e interpretado da seguinte maneira:

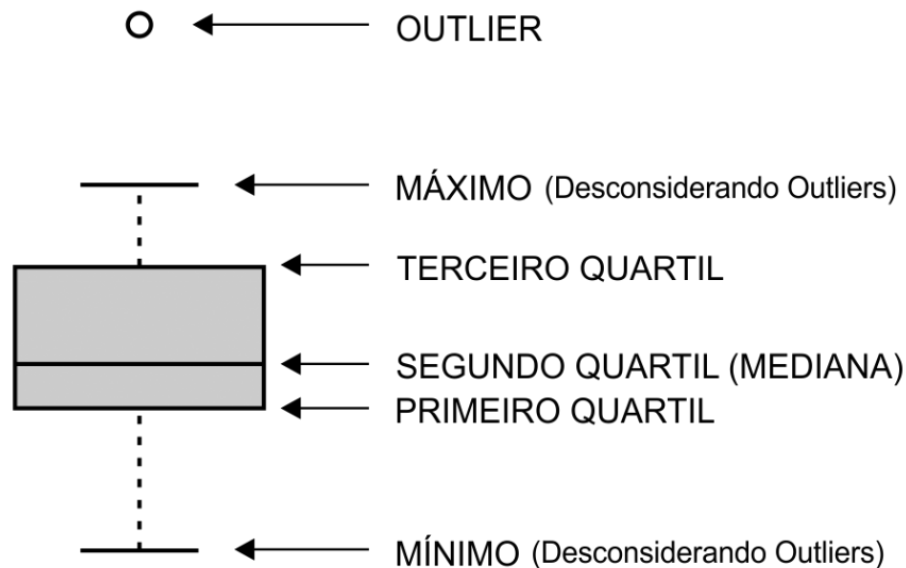


Figura 3.1: BoxPlot - Explicação

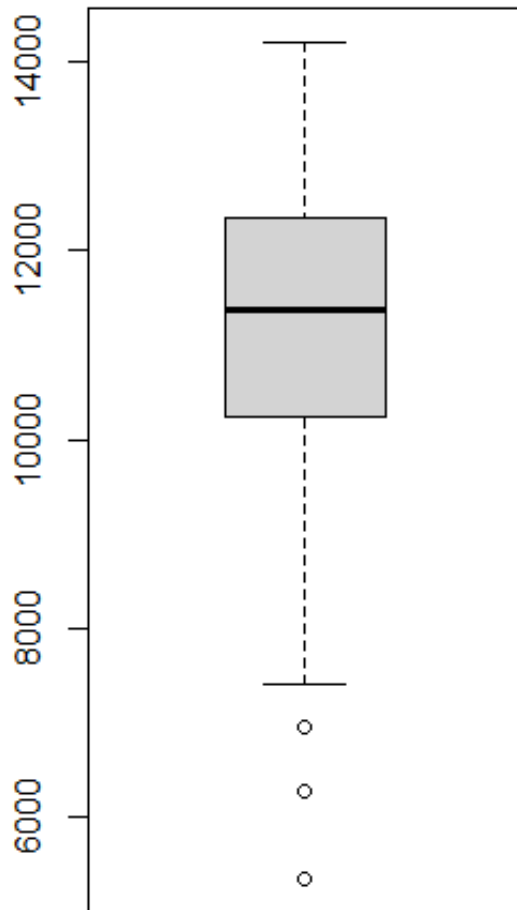
O boxplot (gráfico de caixa) é um gráfico utilizado para avaliar a distribuição empírica dos dados. Tal tipo de gráfico é formado pelo primeiro e terceiro quartil e pela mediana. As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior ao limite inferior e do quartil superior até o maior valor não superior ao limite superior. Para este caso, os pontos fora destes limites são considerados valores discrepantes (outliers).

Sendo assim, com o auxílio da ferramenta R, conseguimos os seguintes BoxPlot's para cada variável analisada anteriormente (Quantidade, Quantidade & Marca e Importação) gerando uma melhor visualização em relação às medidas de tendência central, como explicado anteriormente.

Figura 3.2: Boxplot - Quantidade e Importados

((a))

Boxplot - Quantidade de veículos vendidos



((b))

Boxplot - Quantidade de veículos importados

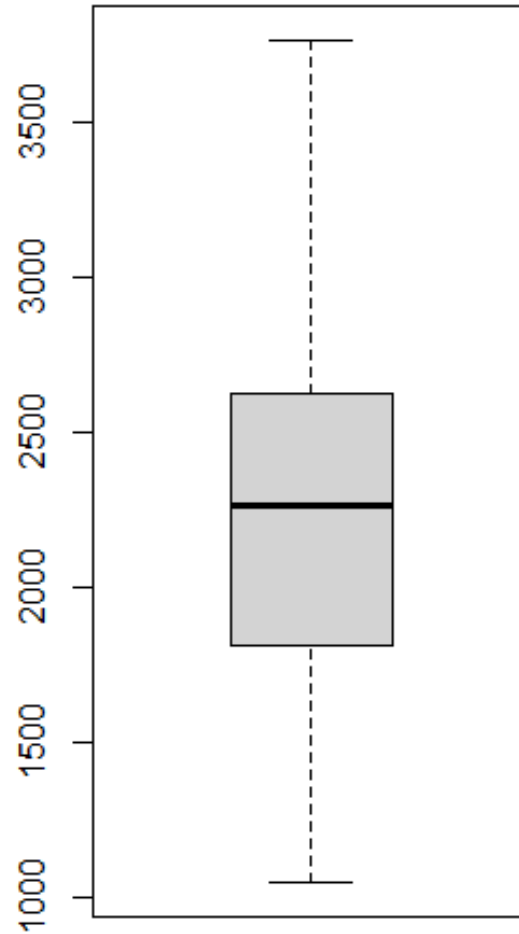
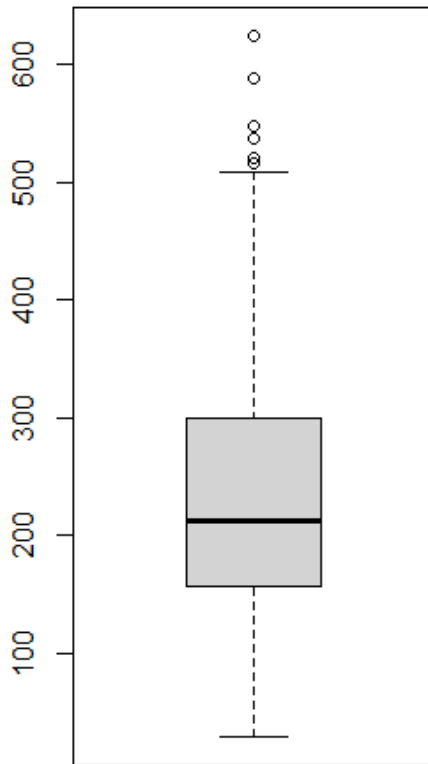
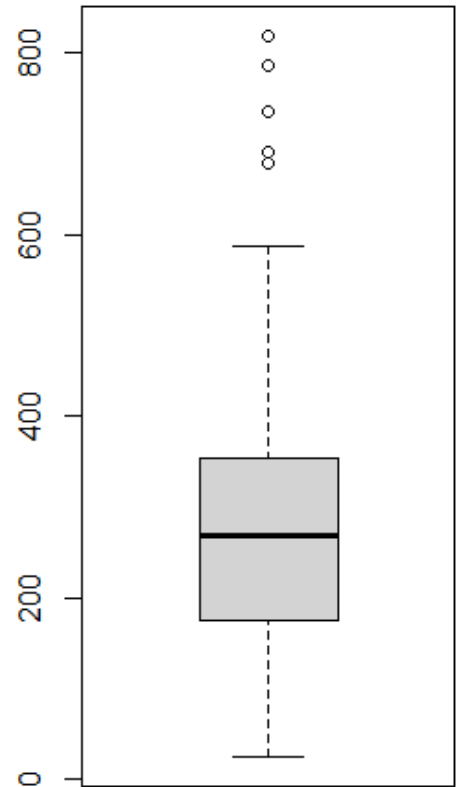


Figura 3.3: Boxplot - Ford e Toyota

((a))

Boxplot - Quantidade de veículos vendidos
Ford

((b))

Boxplot - Quantidade de veículos vendidos
Toyota

3.2 Medidas de dispersão

As medidas de dispersão são medidas que representam a variação dos dados numa base de dados, algumas dessas medidas são: Variância e Desvio Padrão.

Variância: A medida de dispersão denominada Variância tem a função de demonstrar o quão distante cada valor da base de dados está distante do valor médio. Por exemplo, quanto menor a variância, mais perto tais dados estão da medida central (média).

Desvio Padrão: O desvio padrão é capaz de identificar o “erro” em um conjunto de dados, caso quiséssemos substituir um dos valores coletados pela média aritmética.

Medidas de dispersão são importantes para futuramente trabalhar juntamente com o conceito de correlação entre duas variáveis numéricas contínuas. Outro ponto importante para esse tipo de medida estatística é por se demonstrar eficaz na análise da distribuição dos dados.

3.2.1 Medidas de dispersão - Quantidade

A seguir será demonstrado o resultado das medidas de dispersão da variável *Quantidade* que se refere à quantidade de veículos vendidos no período do ano de 2007 até janeiro de 2017.

- **Variância:** A variância, em relação à média, de quantidade de vendas de veículos na Noruega no período de ano 2007 até janeiro de 2017 é de 3069667.
- **Desvio Padrão:** O desvio padrão, em relação à média, de quantidade de vendas de veículos na Noruega no período de ano 2007 até janeiro de 2017 é de 1752 automóveis.

Medidas alcançadas facilmente por meio do seguinte código na linguagem R usando a ferramenta RStudio:

```
var(norway_car_month$Quantity)  
3069667
```

```
sd(norway_car_month$Quantity)  
1752.046
```

3.2.2 Medidas de dispersão - Quantidade & Marca

A seguir será demonstrado o resultado das medidas de dispersão da variável *Quantidade* em relação à variável *Marca* visando a quantidade de veículos vendidos em relação às marcas de veículos *Ford* e *Toyota*.

As seguintes medidas de tendência central para a variável *Quantidade* e *Ford* são:

- **Variância:** A variância, em relação à média, de quantidade de vendas de veículos da **Ford** na Noruega no período de ano 2007 até janeiro de 2017 é de 11102.
- **Desvio Padrão:** O desvio padrão, em relação à média, de quantidade de vendas de veículos da **Ford** na Noruega no período de ano 2007 até janeiro de 2017 é de 105 automóveis.

As seguintes medidas de tendência central para a variável *Quantidade* e *Toyota* são:

- **Variância:** A variância, em relação à média, de quantidade de vendas de veículos da **Toyota** na Noruega no período de ano 2007 até janeiro de 2017 é de 15570.
- **Desvio Padrão:** O desvio padrão, em relação à média, de quantidade de vendas de veículos da **Toyota** na Noruega no período de ano 2007 até janeiro de 2017 é de 124 automóveis.

Medidas alcançadas facilmente por meio do seguinte código na linguagem R usando a ferramenta RStudio:

```
#FORD
var(norway_car_model$Quantity[norway_car_model$Make == 'Ford'])
11102.27
sd(norway_car_model$Quantity[norway_car_model$Make == 'Ford'])
105.3673

#TOYOTA
var(norway_car_model$Quantity[norway_car_model$Make == 'Toyota'])
15570.52
sd(norway_car_model$Quantity[norway_car_model$Make == 'Toyota'])
124.7819
```

3.2.3 Medidas de dispersão - Importações

A seguir será demonstrado o resultado das medidas de dispersão da variável *Importação* que se refere à quantidade de veículos vendidos no período do ano de 2007 até janeiro de 2017.

- **Variância:** A variância, em relação à media, de quantidade de importações de veículos na Noruega no período de ano 2007 até janeiro de 2017 é de 328475.
- **Desvio Padrão:** O desvio padrão, em relação à media, de quantidade de importações de veículos na Noruega no período de ano 2007 até janeiro de 2017 é de 573 automóveis.

Medidas alcançadas facilmente por meio do seguinte código na linguagem R usando a ferramenta RStudio:

```
var(norway_car_month$Import)
328475.6

sd(norway_car_month$Import)
573.1279
```

3.3 Formato da distribuição dos dados

O Formato da distribuição dos dados tem como dever descrever como os dados são distribuídos, explicando então se a distribuição é simétrica ou assimétrica. Existindo três tipos elementares de distribuição de dados para uma variável. Ao analisar as medidas de tendência central, poderemos dizer em qual tipo de formato citado a variável analisada irá se identificar.

será considerado o Índice de Assimetria de Pearson, sendo:

- Entre -0.15 e/ou 0.15: Praticamente Simétrica
- Entre -1 até -0.15 e/ou 0.15 e 1: Assimetria moderada
- Entre Menor que -1 e/ou Maior que 1: Assimetria forte

Também existem análises para verificar pra qual lado está a assimetria, caso ocorra:

- Assimétrica Negativa ou à Esquerda: Média < Mediana < Moda
- Simétrica: Média = Mediana = Moda
- Assimétrica Positiva ou à Direita: Média > Mediana > Moda

É possível indicar o nível de achatamento da curva de distribuição dos dados, ou então, o grau de Curtose da curva. Existindo três tipos elementares: *Leptocúrtica*, *Mesocúrtica* e *Platicúrtica*. Aplicando ferramentas obtidas no software R, pode-se inferir as possíveis análises:

- Aproximadamente 3: Mesocúrtica
- Menor que 3: Platicúrtica
- Maior que 3: Leptocúrtica

Será analisado nas variáveis anteriores (Quantidade, Quantidade & Marca, Importações) os tipos de distribuição dos dados que possuem e o graus de curtose da curva gerada, com auxílio da ferramenta R e RStudio usando os seguintes comandos:

```
install.packages("moments")
library(moments)
skewness(norway_car_month$) #Para o calculo de Assimetria.
kurtosis(norway_car_month$) #Para o calculo de Curtose.
```

Junto de funções encontradas no site universitário UFPR ¹ para visualização de dados, onde o código, e demais outro, para visualização de dados se encontra:

UFPR - Visualização de dados

¹www.leg.ufpr.br/walmes/cursoR/data-vis/01-vis-graphics.html

3.3.1 Formato da distribuição dos dados - Quantidade

Como explicado anteriormente, o formato da curva de distribuição de dados é resultado da análise das medidas de tendência central para descobrir se a distribuição é simétrica ou assimétrica. Sendo assim:

- Média: 11134
- Moda: 11469
- Mediana: 11385

Dessa maneira, compreende-se que: $\text{Moda} > \text{Mediana} > \text{Média}$. E usando a ferramenta R, é facilmente alcançada o formato da distribuição:

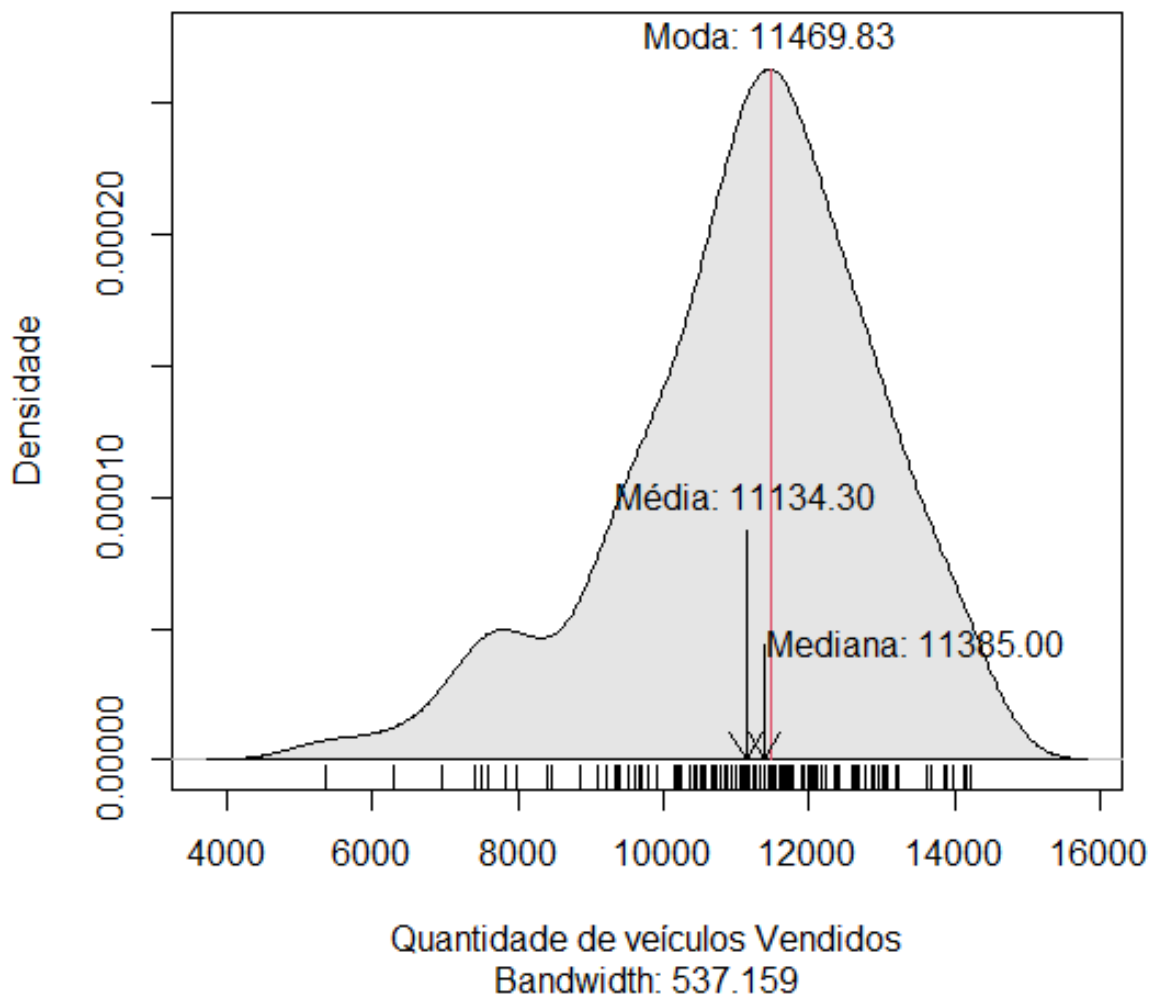


Figura 3.4: Distribuição quantidade de veículos vendidos

Portanto, percebe-se que a distribuição é uma curva de natureza *Assimétrica Moderada* de -0.7349411, porém como a moda é maior que as demais medidas, existe *Assimetria à Esquerda* ou *Assimetria Negativa* nos dados. O grau de Curtose é de natureza *Leptocúrtica* de 3.56929 para a variável *Quantidade* que representa a quantidade de veículos vendidos no período de ano 2007 até janeiro de 2017.

3.3.2 Formato da distribuição dos dados - Quantidade & Marca

Como explicado anteriormente, o formato da curva de distribuição de dados é resultado da análise das medidas de tendência central para descobrir se a distribuição é simétrica ou assimétrica. Sendo assim:

As seguintes medidas de tendência central para a variável *Quantidade* e *Ford* são:

- Média: 234
- Moda: 169
- Mediana: 212

Sendo assim, compreende-se que: Média > Mediana > Moda. E usando a ferramenta R, é facilmente alcançada o formato da distribuição:

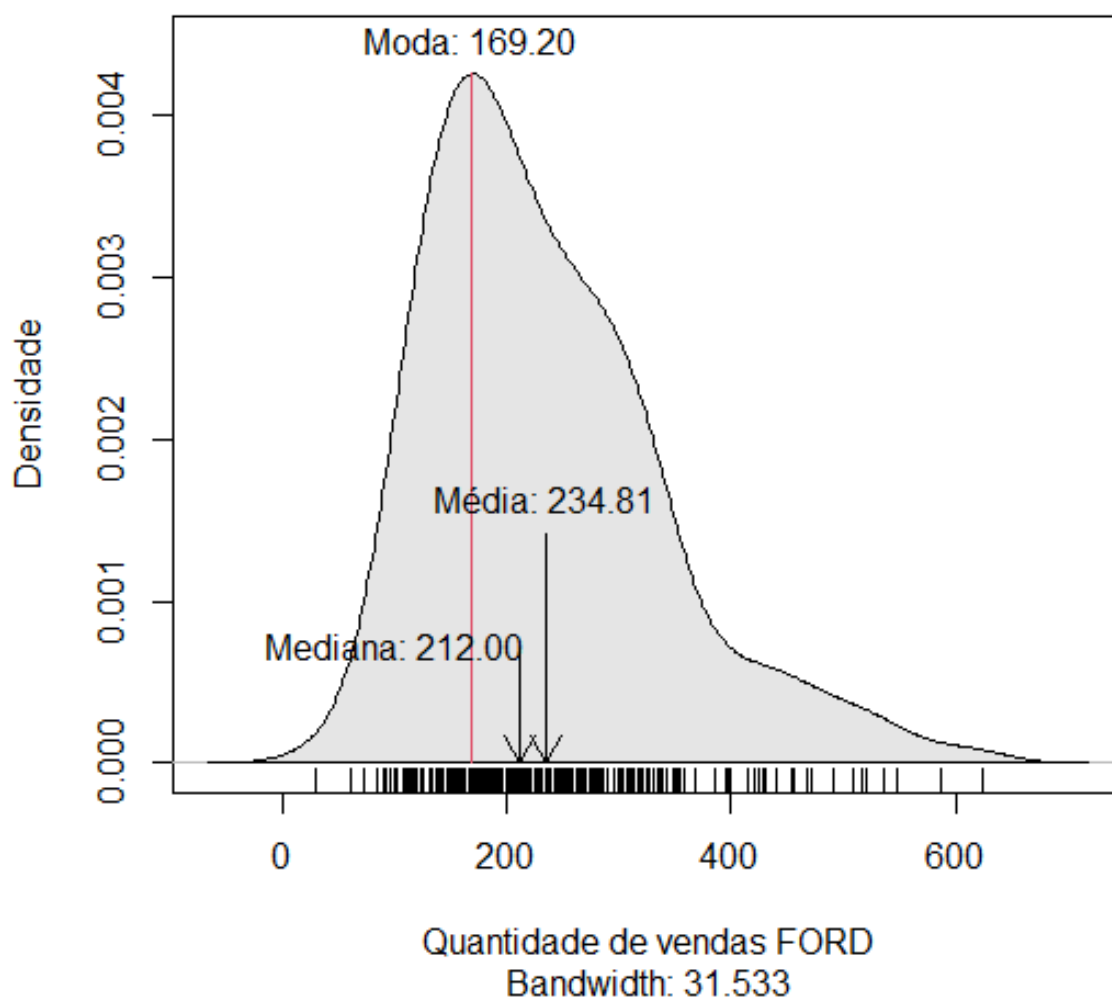


Figura 3.5: Distribuição quantidade de vendas FORD

Portanto, percebe-se que a distribuição é da natureza *Assimétrica Forte* de 1.019903, porém como a média é maior que a mediana e a moda, existe *Assimétrica à Direita* ou *Assimetria Positiva* nos dados. O grau de Curtose é de natureza *Leptocúrtica* de 4.035191 para a variável *Quantidade* em relação à marca *Ford*, que representa a quantidade de veículos vendidos no período de ano 2007 até janeiro de 2017.

As seguintes medidas de tendência central para a variável *Quantidade* e *Toyota* são:

- Média: 273
- Moda: 179
- Mediana: 268

Sendo assim, compreende-se que: Média > Mediana > Moda. E usando a ferramenta R, é facilmente alcançada o formato da distribuição:

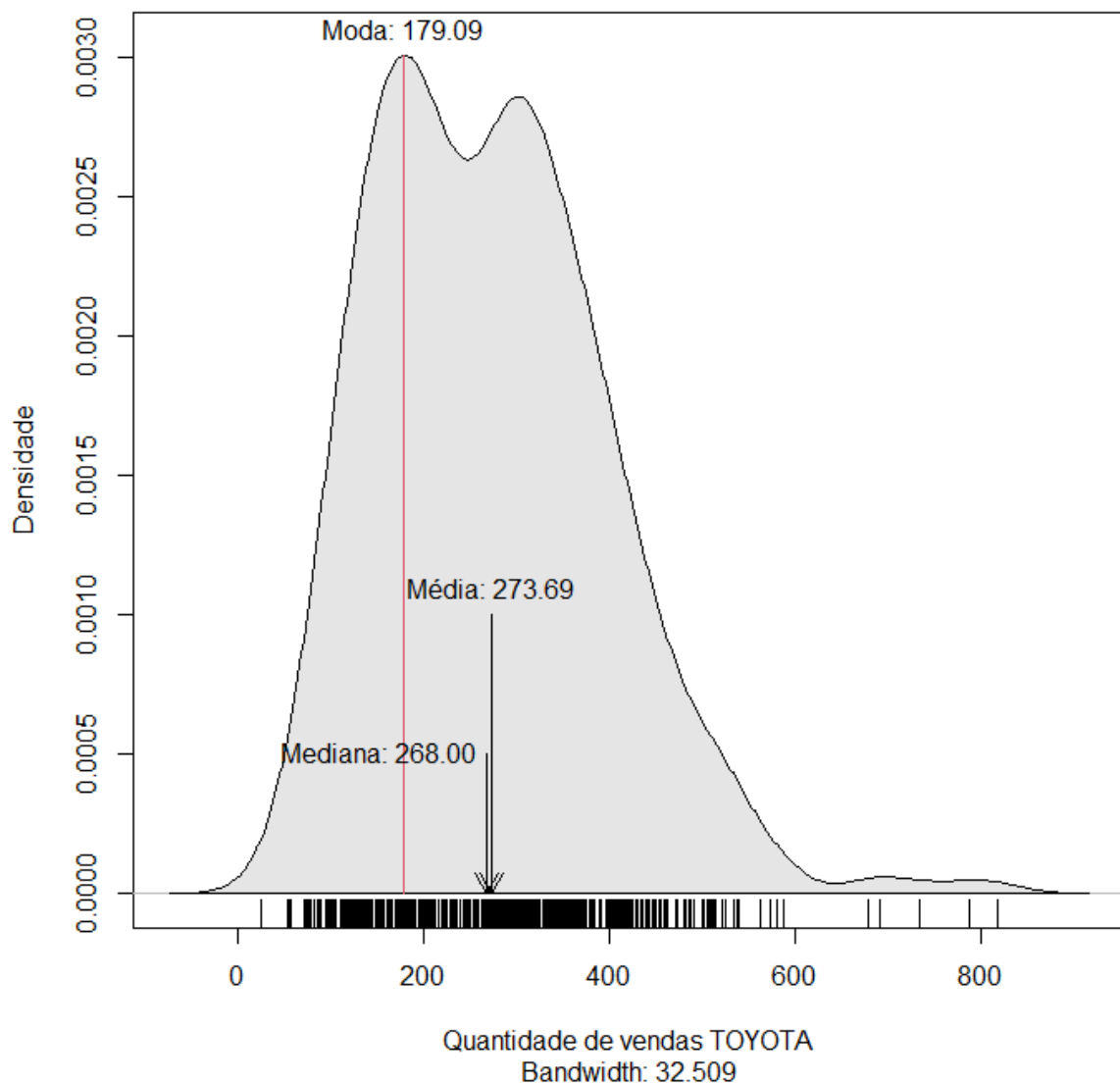


Figura 3.6: Distribuição quantidade de vendas TOYOTA

Portanto percebe-se que a distribuição é da natureza *Assimétrica Moderada* de 0.7423227, porém como a média é maior que a mediana e a moda, existe *Assimétrica Positiva* ou *Assimetria à Direita* nos dados. O grau de Curtose é de natureza *Leptocúrtica* de 3.96476 para a variável *Quantidade* em relação à marca *Toyota*, que representa a quantidade de veículos vendidos no período de ano 2007 até janeiro de 2017.

3.3.3 Formato da distribuição dos dados - Importações

Como explicado anteriormente, o formato da curva de distribuição de dados é resultado da análise das medidas de tendência central para descobrir se a distribuição é simétrica ou assimétrica. Sendo assim:

- Média: 2204
- Moda: 2511
- Mediana: 2263

Sendo assim, compreende-se que: $\text{Moda} > \text{Mediana} > \text{Média}$. E usando a ferramenta R, é facilmente alcançada o formato da distribuição:

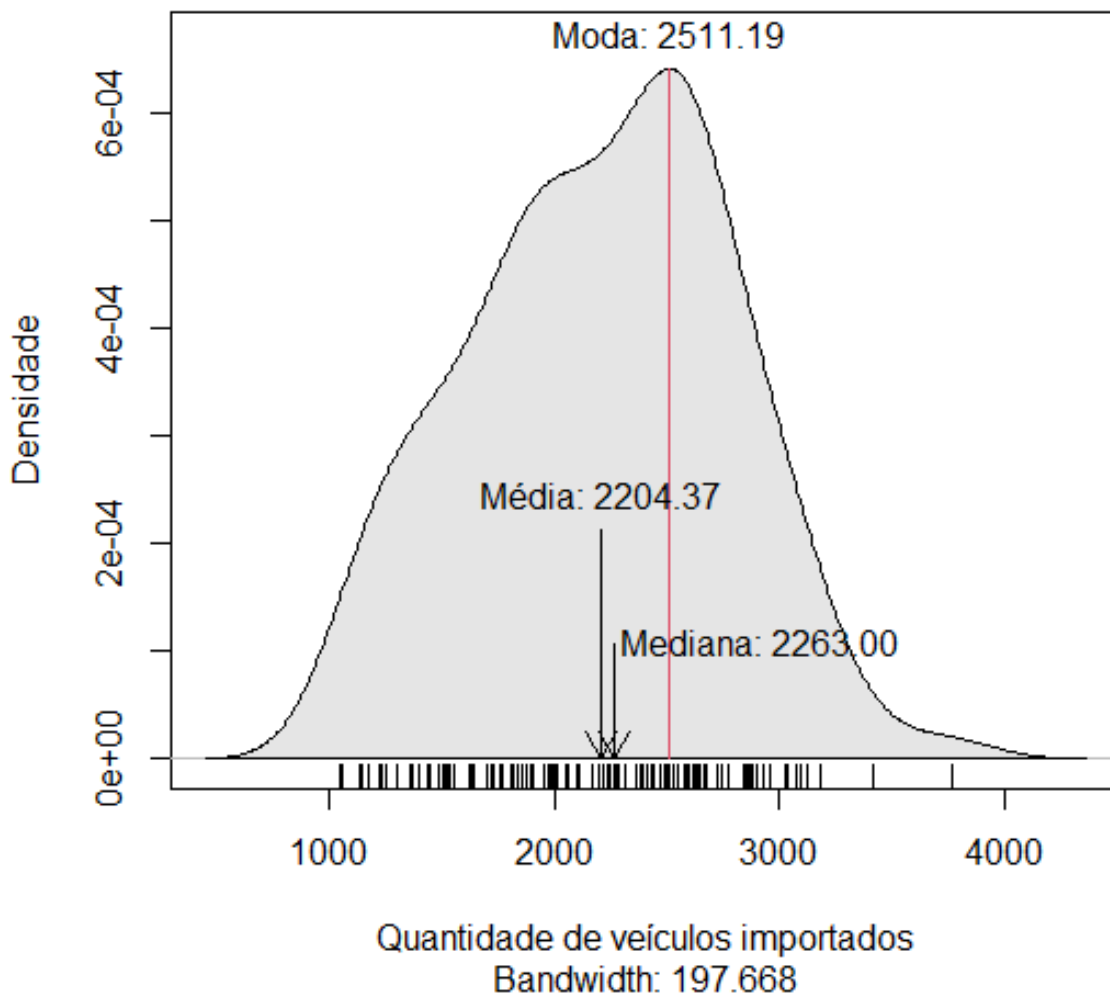


Figura 3.7: Distribuição quantidade de veículos Importados

Portanto percebe-se que a distribuição é uma curva de natureza aproximadamente *Simétrica* de -0.05524526, porém como a moda é maior que as demais medidas, existe *Assimetria à Esquerda* ou *Assimetria Negativa* nos dados. O grau de Curtose é de natureza *Platicúrtica* de 2.443223 para a variável *Importação* que representa a quantidade de veículos importados no período de ano 2007 até janeiro de 2017.

3.4 Correlação entre as variáveis quantitativas

Quando existem variáveis quantitativas, calcula-se a correlação entre elas, podendo variar de 1 até -1, sendo 1 quando as variáveis possuem correlação positiva, -1 quando as variáveis possuem correlação negativa e 0 quando a correlação entre as variáveis é neutra. Com o uso de procedimentos analíticos e gráficos, é possível tornar a análise de correlação mais refinada e precisa.

3.4.1 Ideia geral - Exemplo: Quantidade & Quantidade Diesel

Uma simples análise com as seguintes variáveis: *Quantidade* e *Quantidade Diesel*, pode-se verificar se a Quantidade de carros vendidos no geral possui correlação com a quantidade de carros movidos a diesel vendidos.

Por meio de recursos gráficos na ferramenta R, é gerado o gráfico das duas variáveis, sendo a Quantidade de carros vendidos no geral no Eixo Y, e a Quantidade de carros movidos a diesel vendidos no eixo X.

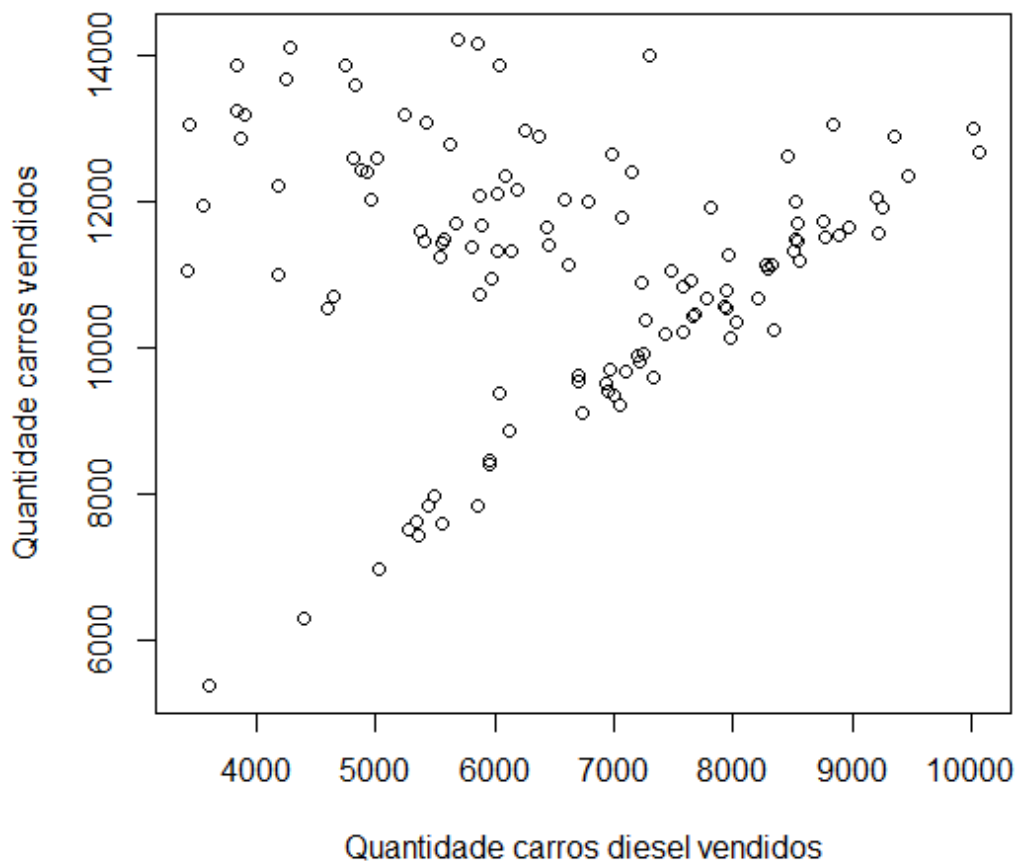


Figura 3.8: Quantidade vs. Quantidade Diesel

Porém ao aplicar a técnica *abline* para plotar uma linha reta, da qual essa segue a ideia de uma regressão, gera-se o seguinte gráfico:

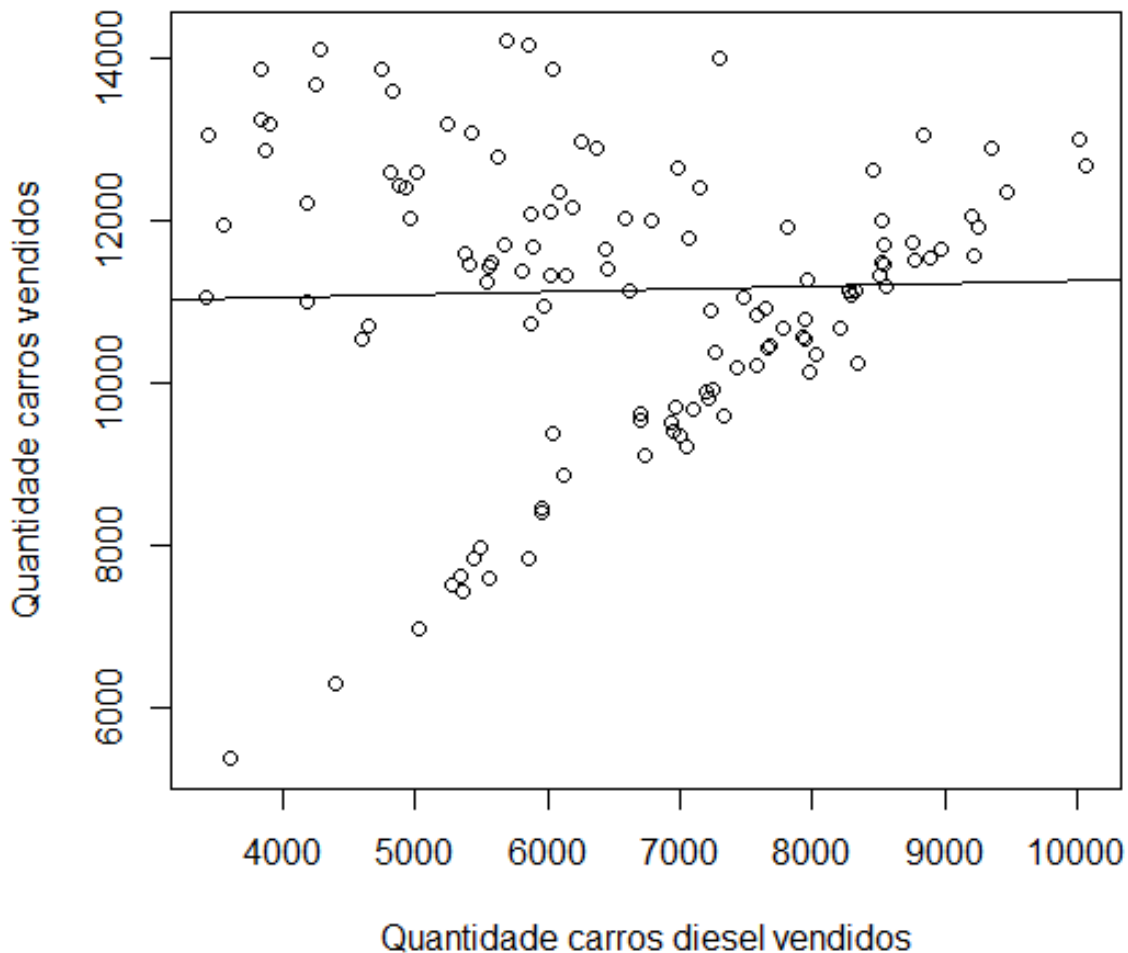


Figura 3.9: Regressão - Quantidade vs. Quantidade Diesel

Após a aplicação da técnica *abline*, é evidente que a correlação entre as duas variáveis pode ser neutra visto que a reta encontrada, de regressão linear, está quase paralela ao eixo das abcissas, o que nos geraria um coeficiente de correlação próximo de 0. Entretanto é necessária análises mais profundas para confirmar tal evidência. Ou seja, a quantidade de carros gerais pode possuir correlação neutra em relação à variável quantidade diesel. Sendo assim, de maneira extremamente rasa, pode-se afirmar que, possivelmente, as variáveis não possuem interferência uma na outra.

3.4.2 Correlação das variáveis - Geral

Visando efetuar a análise de correlação para todas as outras variáveis no banco de dados, das mais distintas maneiras, pôde ser efetuado, com o auxílio da ferramenta de software R e RStudio, a seguinte imagem:

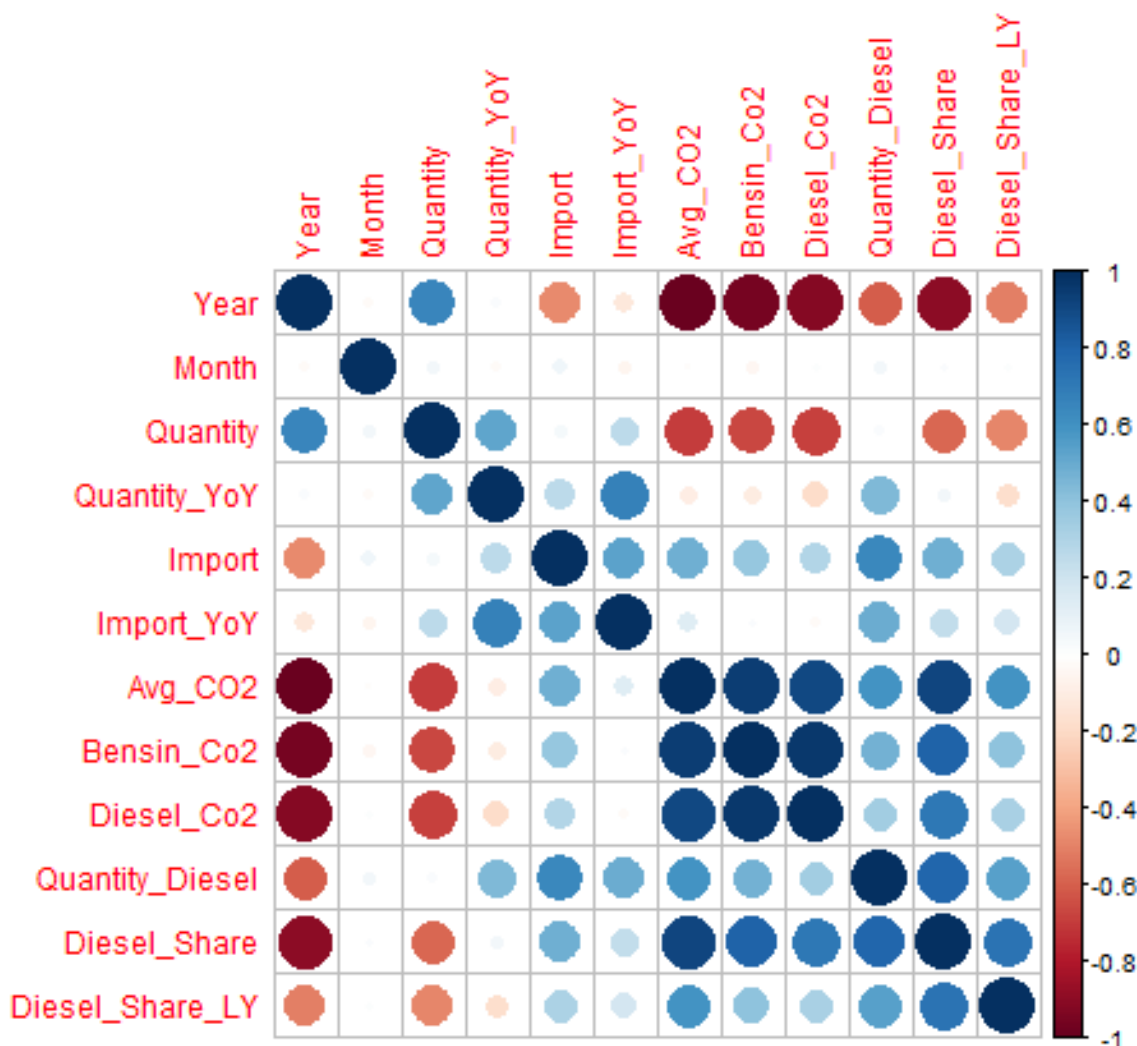


Figura 3.10: Matriz correlação círculo

Com a legenda devidamente posta no gráfico, quanto mais azul o encontro das variáveis na tabela, mais correlação positiva elas possuem. E quanto mais vermelho o encontro das variáveis na tabela, mais correlação negativa elas possuem. E quando branco significa que a correlação entre as possíveis variáveis é neutra.

Dessa maneira, percebe-se que variáveis como Quantidade e Ano possuem forte correlação positiva entre si, ou seja, com o passar dos anos mais carros de maneira geral são vendidos. De maneira análoga, compreende-se que a variável Ano e a quantidade de carros vendidos emitem de CO2 possuem forte correlação negativa, o que nos indica que, ao passar dos anos, menos carros emitem de CO2 são vendidos. Após breve análise de correlação feita, pode-se abrir espaço para confirmar o tema principal deste trabalho.

Para melhor visualização das correlações pode-se efetuar a mesma análise, porém ao invés de círculos, será retornado o real coeficiente de correlação:

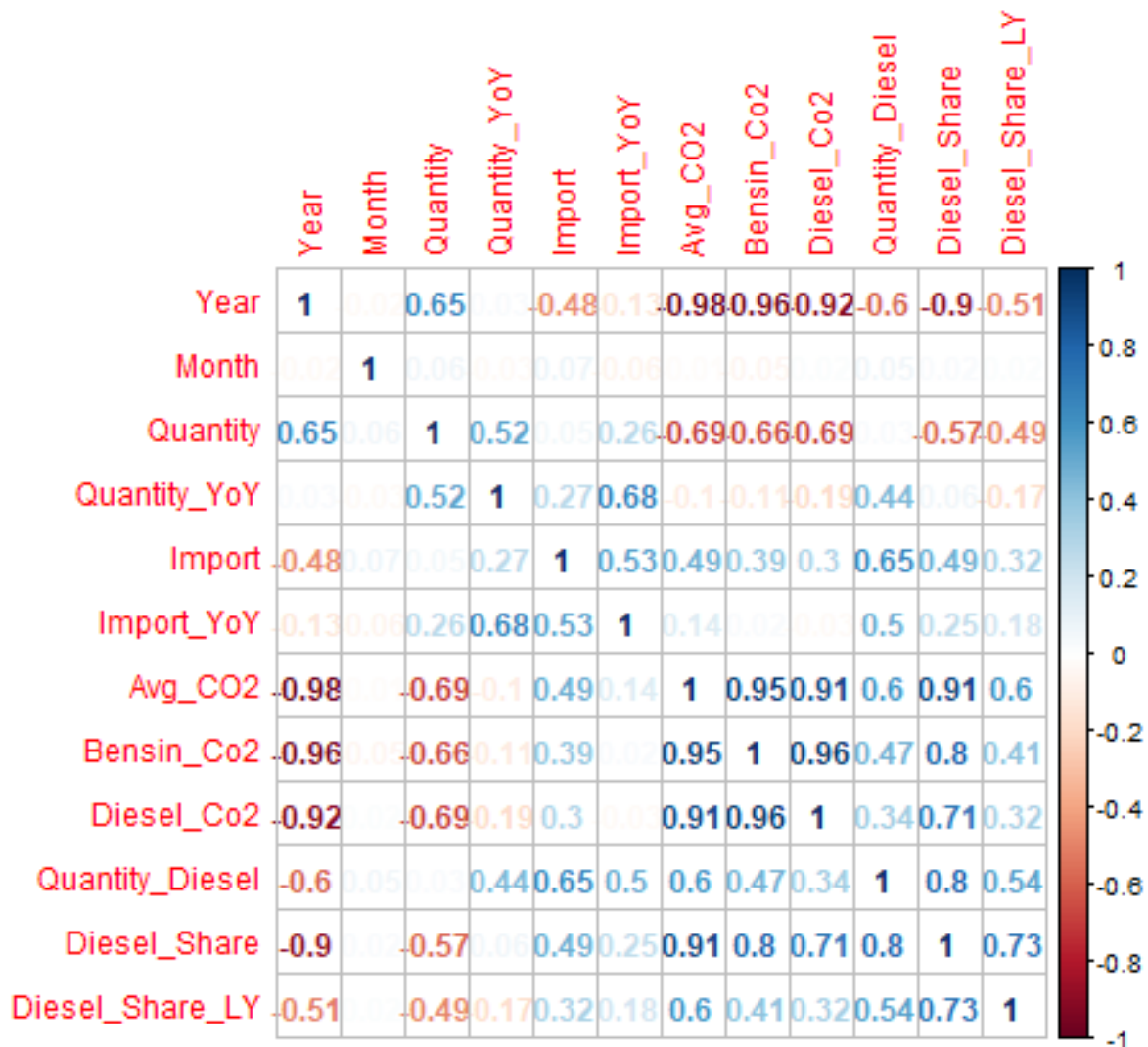


Figura 3.11: Matriz correlação números

Como foi evidenciado anteriormente com a plotagem do gráfico 3.9, as variáveis *Quantidade* e *Quantidade Diesel* possui correlação neutra, pois, como visualizado em nossos diagramas de correlação, o encontro das variáveis é branco, ou seja, 0 ou próximo de 0. Entretanto visualizando as outras variáveis que diz respeito à CO2, como *AVG_CO2*, *Bensin_CO2* e *Diesel_CO2* percebe-se que o encontro dessas variáveis com a variável *Quantidade* possuem correlação negativa. Já o foco deste trabalho, ou seja, a quantidade de carros vendidos emissores de CO2 (*Quantidade Diesel*) em relação ao ano possui correlação negativa equivalente à 0.6.

Pode-se agrupar tais variáveis de maneira que fique mais fácil a visualização, com o método de Cluster. O método Cluster tenta encontrar padrões e agrupa-os de acordo com o padrão encontrado. Sendo assim, após aplicado o Cluster, foi gerado o seguinte resultando:

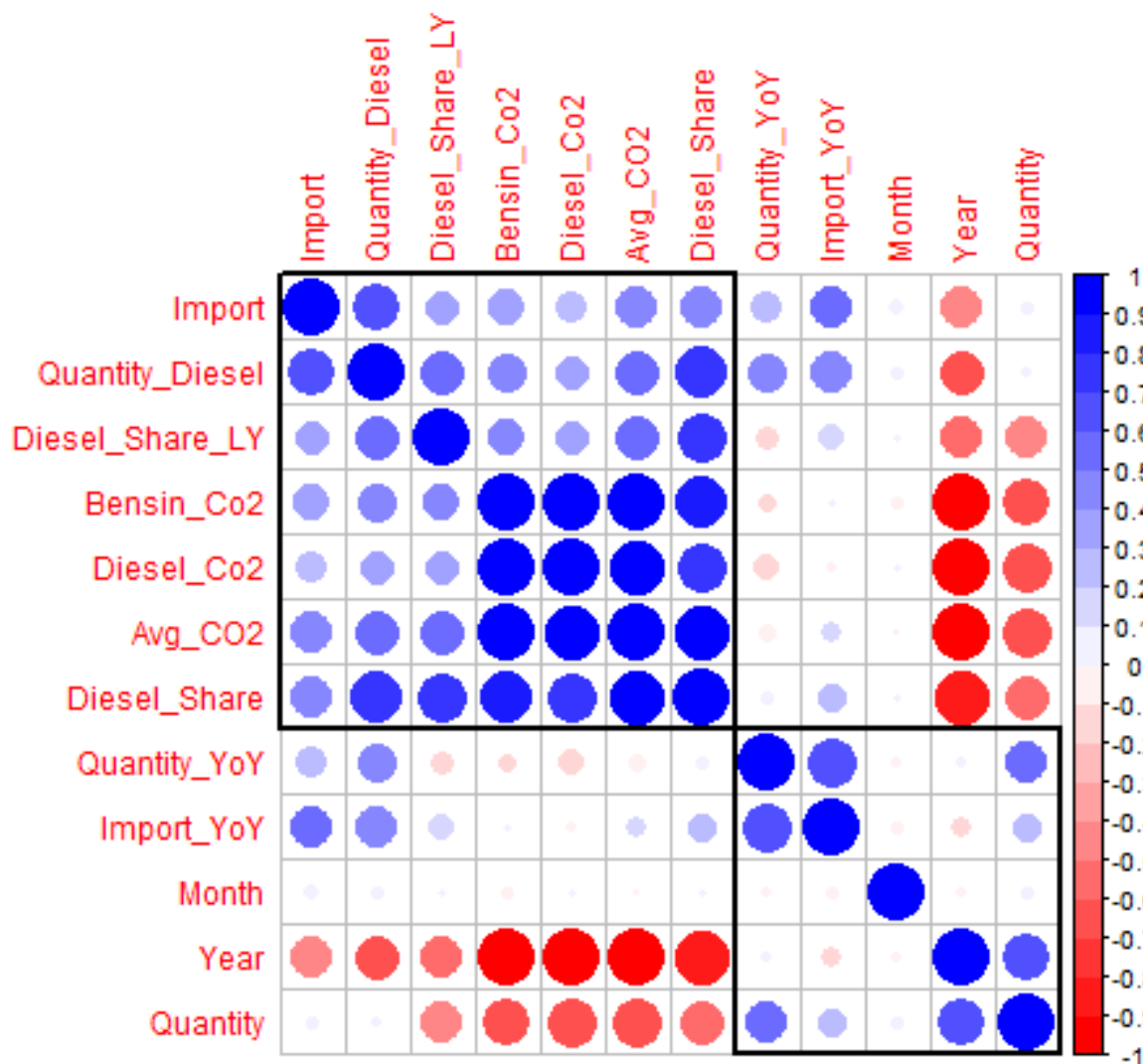


Figura 3.12: Matriz correlação números

Dessa maneira, é evidenciado o resultado de um agrupamento por padrão encontrado nas variáveis, sendo o agrupamento por correlação positiva, negativa ou neutra entre si em relação às variáveis, do qual facilmente é percebido na figura 3.12.

Dessa mesma maneira pode-se fazer diversas análises de maneira simples em relação à correlação de variáveis com o software R e RStudio. Por exemplo, é demonstrado na imagem 3.11, que as variáveis *Quantidade* e *Ano* possuem um grau de correlação positiva de 0.65, tal grau certamente significativo para uma possível análise de negócios pois de maneira rasa, pode-se afirmar que, com o passar dos anos, a quantidade de vendas gerais aumentam.

Capítulo 4

Análise das medidas encontradas

4.1 Introdução às análises

Agora com as análises exploratórias de medidas de tendência central, medidas de dispersão, formato da distribuição e correlação entre as variáveis, pode-se analisar de maneira mais específica cada variável, no qual serão separados subcapítulos, como foi efetuado anteriormente em capítulos passados.

4.1.1 Análise - Quantidade

A quantidade de vendas, seja em qualquer negócio, é um fator muito importante e impactante, principalmente vendas que alavancam um país economicamente, como é o caso dessa base de dados do país nórdico, Noruega.

Como encontrado no capítulo 3, a média, mediana, variância, desvio padrão, formato de distribuição dos dados e correlação. De acordo com estudos efetuados na nossa base de dados foi possível encontrar as seguintes medidas para a variável *Quantidade*:

- **Média:** 11134 veículos vendidos por mês.
- **Moda:** 11469 veículos vendidos no mês.
- **Mediana:** 11385 veículos vendidos.
- **Variância:** É de 3069667 em relação aos dados de vendas de veículos.
- **Desvio Padrão:** 1752 veículos vendidos.
- **Índice de Assimetria:** Assimetria à Esquerda de -0.7349411.
- **Curtose:** Grau de Curtose de 3.56929.

De acordo com essas medidas pode-se ver que a Média, Moda e a Mediana estão ligeiramente próximas, portanto os dados parecem estar distribuídos de maneira simétrica, porém, como foi analisado no capítulo 3.3.1 é evidenciado que existe uma assimetria moderada. Na distribuição dos dados também percebe-se um grande Grau de Curtose, que nos conclui a distribuição de uma curva Leptocúrtica.

Com o cálculo efetuado sobre as medidas de dispersão, é evidenciado grande variação nos dados junto com alta taxa de desvio padrão para a quantidade de vendas de veículos. Porém, um fato a ser levado em consideração (E será melhor evidenciado posteriormente com auxílio de Séries Temporais) é de que, nos anos de 2008 e 2009 houve uma forte crise econômica global, atingindo fortemente a venda de veículos na Noruega nesse período, tornando os dados, talvez, mais variados por esse motivo. Vemos também que tal variável possui correlação com as demais, sendo tais correlações positivas ou negativas, dificilmente são neutras, ou seja, é uma variável de suma importância. Por exemplo, como citado anteriormente que talvez o ano poderia influenciar na quantidade de vendas de veículos (Por causa da crise de 2008/2009, no exemplo citado), analisando a figura 3.11 facilmente é visto que realmente possuem correlação pois o grau de correlação é de 0.65, sendo assim, existe correlação entre o ano e a quantidade de vendas de veículos.

Portanto, é de grande surpresa, após conseguir essas medidas, que, um país com a população de aproximadamente 5 milhões de pessoas, tenha a média de venda de veículos de **onze mil**, ou levando em consideração o desvio padrão para cima: **até treze mil** veículos, mesmo após enfrentar uma grave crise econômica global, o que evidencia que, possivelmente, a população é economicamente ativa em relação aos seus bens materiais.

4.1.2 Análise - Quantidade & Marca

Análise em cima de algumas marcas automobilísticas, como feito anteriormente com as marcas Ford e Toyota. Sendo assim, as medidas das variáveis *Quantidade & Ford* são:

- **Média:** 234 veículos vendidos por mês.
- **Moda:** 169 veículos vendidos por mês.
- **Mediana:** 212 veículos vendidos.
- **Variância:** É de 11102 em relação aos dados de vendas de veículos.
- **Desvio Padrão:** 105 veículos vendidos.
- **Índice de Assimetria:** Assimetria à Direita de 1.019903.
- **Curtose:** Grau de Curtose de 4.035191.

As seguinte medidas foram conseguidas através de análises na variável *Quantidade & Toyota*:

- **Média:** 273 veículos vendidos por mês.
- **Moda:** 179 veículos vendidos por mês.
- **Mediana:** 268 veículos vendidos.
- **Variância:** É de 15570 em relação aos dados de vendas de veículos.
- **Desvio Padrão:** 124 veículos vendidos.
- **Índice de Assimetria:** Assimetria à Direita de 0.7423227.
- **Curtose:** Grau de Curtose de 3.96476.

Após a conclusão de todas as medidas para os dois tipos de marca, conclui-se que, ligeiramente, a marca *Toyota* obteve melhores resultados em quesito de venda de automóveis em relação à outra marca analisada, *Ford*. Sendo assim, talvez a marca *Ford* deveria rever a estratégia de vendas nesse país para alavancá-los e triunfar sobre um de seus principais concorrentes, fazendo isto com números expressivamente melhores do que foram encontrados nas análises feitas desse trabalho durante o ano de 2007 até janeiro de 2017. Certamente tal investimento em estratégia de venda faria muito sentido, pois como visto anteriormente, a Noruega é um país em que a compra de carros pela população normalmente cresce ao passar dos anos, e com certeza é um mercado do qual vale a pena o investimento para dominar o mercado de vendas automobilísticas. Após análises de regressões lineares será evidenciado possivelmente um novo motivo por a marca *Ford* não ter se saído tão bem quanto a concorrente *Toyota* o fez.

4.1.3 Análise - Importação

A Noruega, de acordo com o site **CEIC - Importações Geral Noruega** ¹, disponível em: CEIC - Importações Geral Noruega.

Possui um avançado mercado de importação, podendo verificar que o país é uma nação do qual a parte econômica de importação (No geral) é muito ativa e comumente em alta. Visando isso, apresentamos as medidas encontradas para a variável **Importação** nos anos de 2007 até janeiro de 2017:

- **Média:** 2204 veículos importados por mês.
- **Moda:** 2511 veículos importados por mês.
- **Mediana:** 2263 veículos importados.
- **Variância:** É de 328475 em relação aos dados de vendas de veículos.
- **Desvio Padrão:** 573 veículos vendidos.
- **Índice de Assimetria:** Simétrica de -0.05524526.
- **Curtose:** Grau de Curtose de 2.443223.

Como é revelado pelo portal CEIC, percebe-se que a taxa geral de importação é positiva durante esse período de dez anos que foram analisados nesse trabalho, sendo então, a Noruega, um país que realmente trabalha com importações em seu setor econômico, e possivelmente também, no setor econômico automobilístico. Além disso, como visto na imagem 3.7, vemos que o grau de assimetria é de -0.05, que, de acordo com o Índice de assimetria de Pearson, pode-se considerar a distribuição simétrica. Então, é indicado que a quantidade de importação não foi tão alta quanto poderia ser, porém foi bem concentrada, o que é evidenciado também na média, moda e mediana, que são bem próximas, e como é visualizado também na figura 3.2(b) do qual possui a representação das medidas de tendência central no boxplot para a variável *Importação*.

Além disso, é possível verificar na figura 3.11 que a variável *Import* possui correlação positiva com todas as outras variáveis, exceto com a variável *Year* da qual possui correlação negativa. Portanto conclui-se que, ao passar dos anos a quantidade de importação de veículos efetuada pela Noruega, desce.

¹www.ceicdata.com/pt/indicator/norway/total-imports

Capítulo 5

Regressões e Previsões

5.1 Regressão Linear

O conceito de regressão tenta ajudar a prever comportamentos com base na associação entre duas variáveis numéricas contínuas que geralmente possuem alguma correlação. Existem diversos tipos de regressões, porém nesse trabalho será utilizado apenas a regressão de tipo Linear. A Regressão Linear, como seu próprio nome diz, trabalha com funções lineares e tentará confeccionar uma reta que passa o mais próximo de cada ponto existente no plano cartesiano após a plotagem das duas variáveis em cada eixo cartesiano.

Nesse capítulo será aprofundado a ideia principal desse trabalho, sendo essa tentar provar que a venda de carros elétricos e/ou híbridos está crescendo, analisando apenas a venda dos tipos de carro tradicionais, ou seja, avaliando os carros que emitem CO₂ (Quantity_Diesel) e as vendas de carros gerais (Quantity). Por meio de regressões e análises de correlação, juntamente com as análises já feitas no capítulo 3 na figura 3.11 será feita uma investigação para confirmar que a venda de carros elétricos e/ou híbridos realmente vêm crescendo e sendo majoritária em relação à venda geral de carros nos passar dos anos.

Sendo assim, por meio de gráficos nas ferramentas R e RStudio foi possível concretizar regressões lineares sobre as variáveis citadas, dos quais geraram conclusões sobre o tema principal deste trabalho.

5.2 Regressão Linear - CO2 vendidos vs. Ano

Como foi citado anteriormente no capítulo 1, a Noruega nos dias atuais se tornou uma grande potência no quesito de vendas de carros elétricos, porém como a base de dados é composta de dados até a data de janeiro do ano de 2017, não pôde ser feito tal análise pois possuíamos muitos dados faltantes em relação ao total, pois os dados de carros elétricos e/ou híbridos começaram a ser introduzidos apenas a partir de 2011, 4 anos após o início da coleta de dados dessa base.

Então, uma maneira de verificarmos o crescimento da venda de carros elétricos, seria analisarmos o contrário, ou seja, carros tradicionais, dos quais são movidos à CO2 por ano. Sendo assim, analisaremos carros vendidos que possuem emissão de CO2 ao longo dos anos:

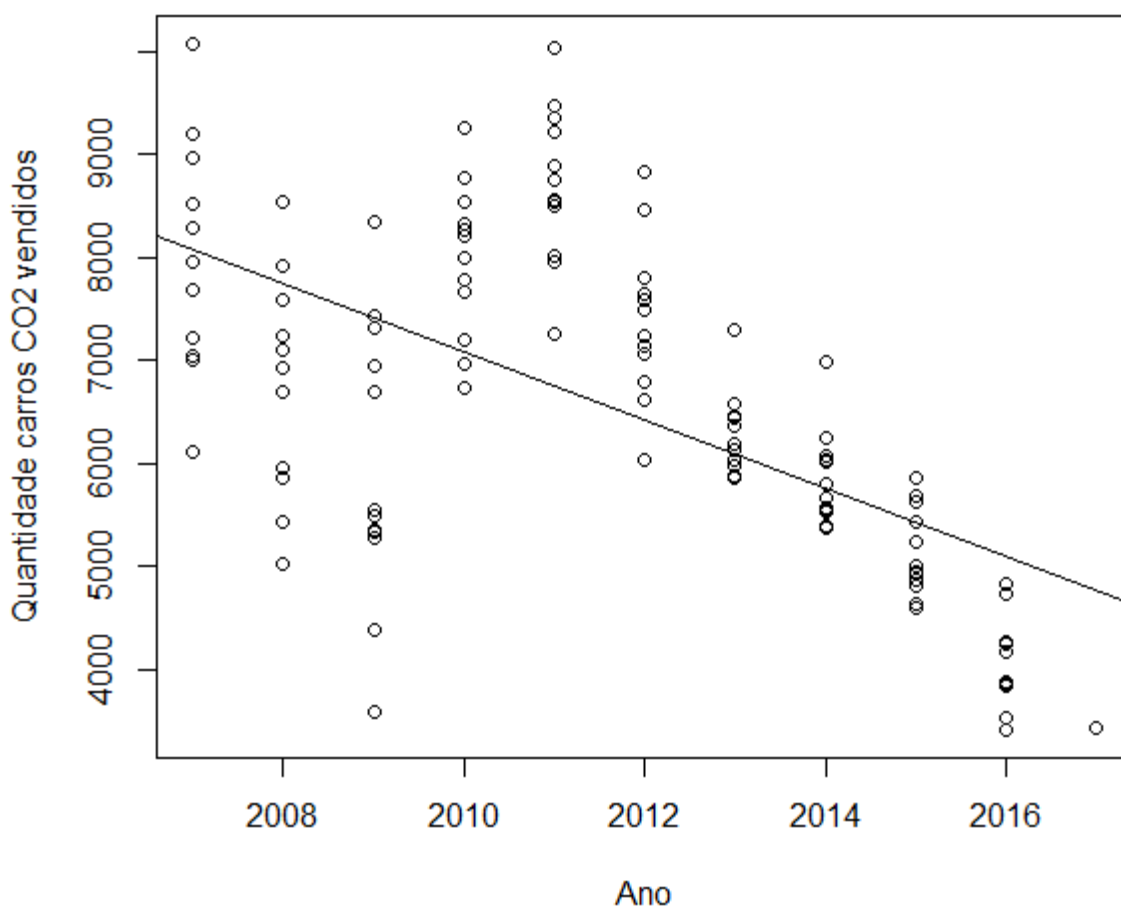


Figura 5.1: Regressão Linear - CO2 vendidos vs. Ano

Com a regressão finalizada sobre as duas variáveis (Quantity_Diesel e Year) podemos verificar que realmente, ao passar dos anos, ou especificamente a partir de 2011, houve queda drástica e contínua em relação aos carros vendidos que emitem CO2.

5.3 Regressão Linear - Quantidade vendidos vs. Ano

Em contra partida analisaremos agora a quantidade de carros (De forma geral) vendidos por ano, tentando verificar se a venda de carros gerais aumentou ao passar dos anos. Se sim, nos mostra que a população está, possivelmente, abandonando cada vez mais o modelo de carro tradicional emissor de CO₂ e migrando para modelos elétricos e/ou híbridos (Pois como visto na figura 5.1 existe queda da venda de carros emitente de CO₂ ao passar dos anos). Portanto, por meio das variáveis *Quantity* e *Year*, foi gerado o seguinte gráfico de regressão linear:

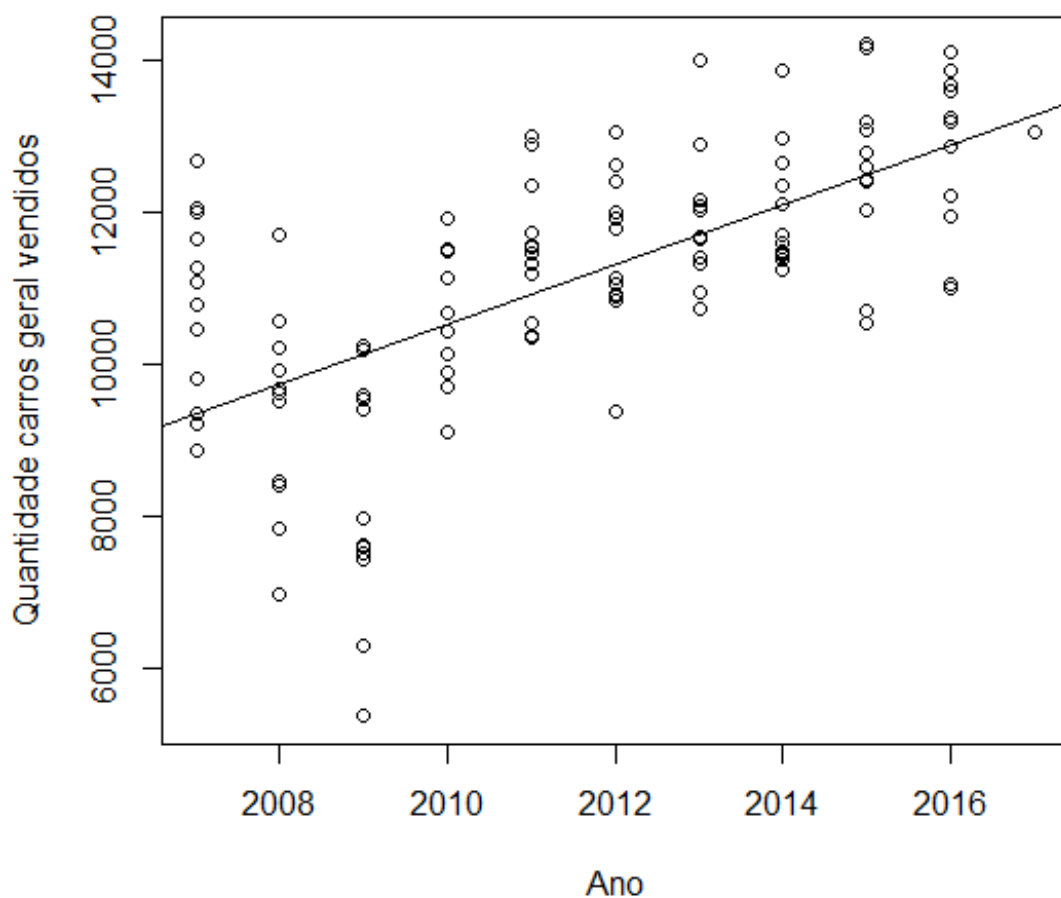


Figura 5.2: Regressão Linear - Quantidade geral veículos vendidos vs. Ano

Como era esperado após análises anteriores, a Quantidade de carros no geral aumentou ao passar dos anos. E outro fator interessante, pôde ser visto no período de crise global, nos anos de 2008 e 2009, o reflexo da análise feita no capítulo 4, do qual houve grande queda na quantidade de carros vendidos, ou seja, talvez por essa grande queda nesses dois anos os dados das variáveis podem ter se tornado bastante variados entre si. Entretanto, percebe-se que, em um curto período a quantidade de carros vendidos em geral voltou a crescer de forma contínua na Noruega.

Portanto, pode-se dizer que a quantidade de carros emissores de CO₂ diminuiu, porém em contra partida a quantidade de carros no geral aumentou com o passar dos anos, nos levando à ideia referenciada pela revista EPOCA, de que realmente a Noruega é um país potência em relação às vendas de carros híbridos e/ou elétricos. Isso mostra o porquê a marca *Ford* não possuiu tanto sucesso em relação às vendas de automóveis do que a

marca concorrente, *Toyota*. A marca *Ford* não possuía modelos de carros elétricos a serem vendidos (Nessa data de janeiro de 2017), acabando por não se inserir tão bem quanto a concorrente *Toyota* o fez, pois a marca *Toyota* possui carros elétricos à venda no mercado.

Capítulo 6

Série Temporais - Análise

6.1 Séries Temporais

No ramo dos negócios, comumente é levado em consideração as vendas conquistadas em cada mês para cada ano, descobrindo talvez possíveis datas padrões das quais necessitariam de eventos de marketing mais eficazes para otimizar as vendas. Uma ferramenta que auxilia quanto à isto é o uso de Séries Temporais. Com as Séries temporais, pode-se verificar o as vendas concluídas em cada ano, ou até melhor, as vendas concluídas em cada mês de cada ano, abrindo caminho assim para possíveis previsões.

6.2 Série Temporal - Quantidade vs. Ano

Plotando uma Série Temporal geral, com auxílio da ferramenta R e RStudio, consegue-se a confecção da seguinte imagem:

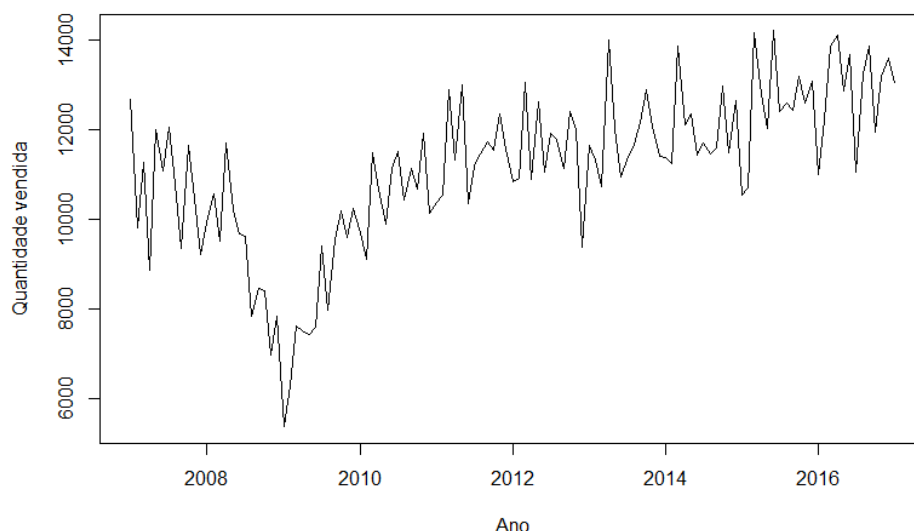


Figura 6.1: Série Temporal - Quantidade vendidos geral vs. Ano

E agora, com toda certeza pode-se inferir, no período de crise global, nos anos de 2008 e 2009, houve uma drástica queda nas vendas de veículos na Noruega por decorrência de tal acontecimento. Percebe-se, inicialmente, sazonalidade nessa Série Temporal, o que, posteriormente, certamente poderá influenciar nos modelos de previsão.

6.2.1 Série Temporal - Quantidade vs. Mês e Ano

Para haver análises em cada mês de cada ano, separa-se da seguinte maneira:

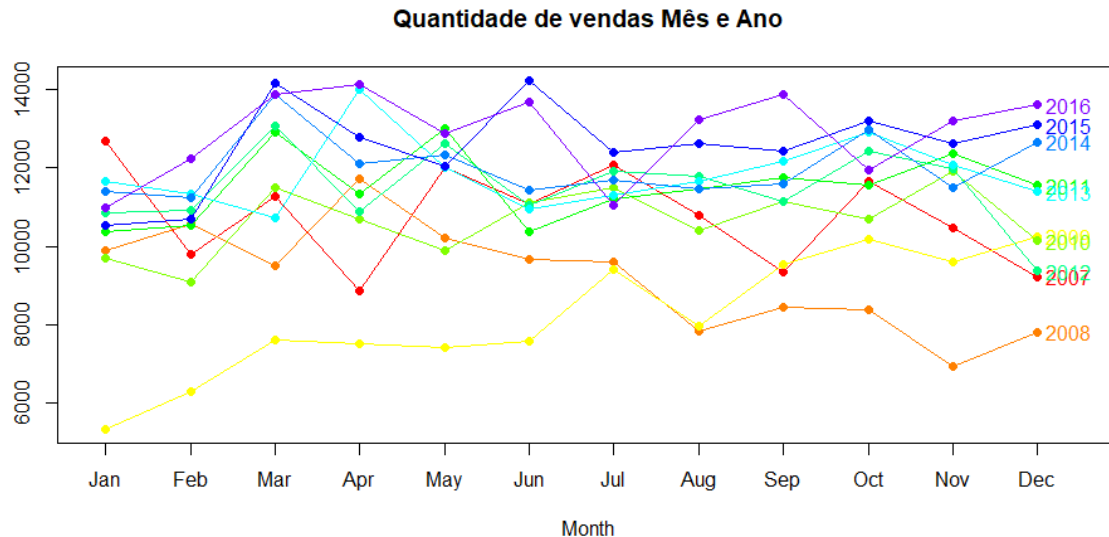


Figura 6.2: Série Temporal - Quantidade vendidos vs. Mês e Ano

Dessa maneira pode-se fazer uma análise cirúrgica para cada ano. Em 2007 por exemplo, houve grande variação nos dados de cada mês, possivelmente uma má fiscalização sobre a variável Quantidade de veículos vendidos, uma vez que a coleta desses dados era iniciada nesse ano. Outro breve exemplo do qual já foi evidenciado, em 2008, no início da crise houve grande queda até o fim do mesmo ano, iniciando 2009 com grande queda e tal ramo apenas pôde se recuperar no segundo semestre de 2009. Revelando então a grande recuperação econômica, após uma crise global, de uma das potências mundiais, como é a Noruega.

Modelos facilmente concluídos com ajuda do software R e RStudio com o seguinte código:

```
install.packages("forecast")
library(forecast)

#Serie Temporal (QUANTIDADE DE CARROS VENDIDOS NOS ANOS)
vetor <- as.numeric(norway_car_month$Quantity)
z <- ts(vetor, frequency = 12, start = c(2007,1),
        end = c(2017,1))

#Quantidade geral de carros vendidos
ts.plot(z)

#Quantidade mes a mes em cada ano de carros vendidos
seasonplot(z, col = rainbow(12), border = c("royalblue"),
           year.labels = TRUE, Type = "o", pch= 16,
           main = "Quantidade_de_vendas_Mes_e_Ano")
```

Capítulo 7

Previsão - Análises

7.1 Previsão

Como mencionado no capítulo 6, após a finalização do modelo temporal e sua respectiva análise, abre-se caminho para efetuar previsões nos próximos meses e/ou anos. Sendo assim, tentaremos prever a quantidade de vendas de carros na Noruega nos próximos trinta e seis meses.

Previsões são sempre bem vindas no âmbito de negócios, porém nesse caso, previsões são extremamente necessárias, pois se refere a uma parcela da economia de um país, o setor automotivo. Nenhuma previsão é certa, pois não se sabe o futuro, do qual esse pode guardar possíveis crises. Entretanto, tentar fazer a previsão em um cenário comum se torna extremamente necessário, pois as marcas que atuam neste país podem tomar decisões melhores com base em possíveis previsões. Tais previsões podem ajudar também o governo a tentar especular o quanto a economia poderá crescer no âmbito automobilístico.

Sendo assim, uma boa previsão é feita após comparar diversos modelos entre si, comparando-as principalmente as acurácias encontradas em cada um. Desta maneira, os modelos dos quais serão utilizados são: Holt-Winters, ARIMA e Naive. Todos modelos de previsão porém possuem seus respectivos detalhes.

Vale ressaltar que a maioria dos modelos de previsão e o uso de bibliotecas como `forecast` no R foram estudadas previamente em: Hyndman e Athanasopoulos (2018).

7.1.1 Previsão com Holt-Winters

O modelo Holt-Winters é considerado um modelo de previsão de suavização exponencial com tendência e sazonalidade. Sendo assim, dentro desse modelo foi-se utilizado artifícios como:

- **Método Aditivo:** Adequado quando a amplitude dos ciclos de sazonalidade não está correlacionada ao tempo.
- **Método Multiplicativo:** Adequado quando a amplitude dos ciclos de sazonalidade está correlacionados ao tempo.

Levando em consideração a figura 6.1, o modelo de previsão Holt Winters previu a distribuição da seguinte maneira:

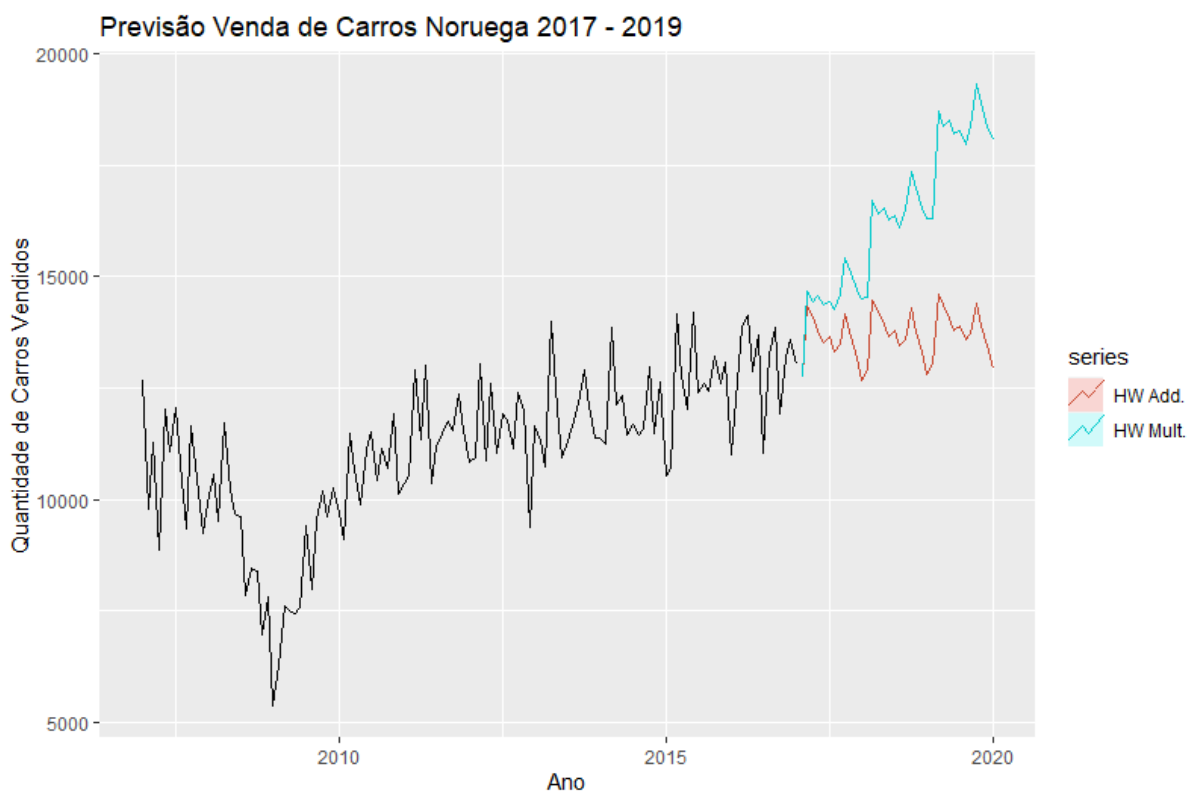


Figura 7.1: Holt Winters - Previsão

Pelo gráfico gerado, existem indícios que o modelo Holt Winters usando o *Método Multiplicativo* se mostrou discrepante em relação ao *Método Aditivo*, sendo assim, para haver a confirmação, basta analisar a acurácia de cada método.

A acurácia do *Método Multiplicativo* se deu por:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	14.44626	1027.328	833.9769	-0.3365531	7.886191	0.6169389	-0.1585527

Figura 7.2: Acurácia - Método Multiplicativo

Já a acurácia do *Método Aditivo* se deu por:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	8.223223	1004.556	794.1782	-0.7223479	7.498917	0.5874976	-0.1661407

Figura 7.3: Acurácia - Método Aditivo

7.1.2 Previsão com ARIMA

O modelo ARIMA é, sem dúvidas, um dos mais conhecidos quando se fala sobre previsão. O nome do modelo é uma sigla, que depois de traduzida significa: Modelo auto-regressivo integrado de médias móveis. Dessa maneira, com o uso da biblioteca *Forecast*, da qual nesta possui funções para estimar modelos ARIMA de modo automático, foi gerada a seguinte distribuição em relação à quantidade de carros vendidos na Noruega:

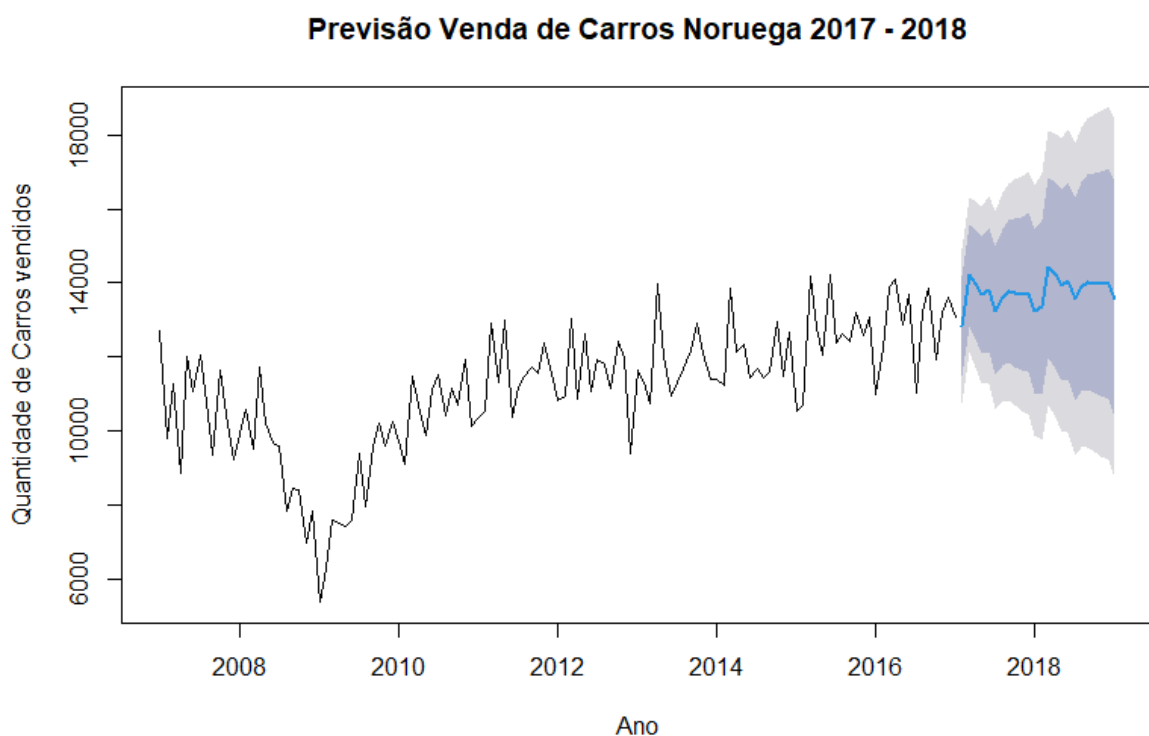


Figura 7.4: ARIMA - Previsão

Dessa maneira, no gráfico gerado existe um seguimento na cor azul claro e outros dois tipos de silhueta. A silhueta cinza representa possíveis valores com 95% de certeza, já a silhueta azul escuro representa possíveis valores com 80% de certeza. Porém, percebemos que tal modelo só foi capaz de prever dados de fevereiro de 2017 até janeiro de 2019.

Avaliando a acurácia do modelo ARIMA:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	34.48725	1036.698	828.5122	-0.4354995	7.892494	0.6128964	0.03260806

Figura 7.5: Acurácia - ARIMA

7.1.3 Previsão com NAIVE

O modelo NAIVE leva em considerações ações que ocorreram no passado da série temporal. Então, tal modelo presume que o futuro repetirá o passado, portanto a previsão corresponde ao último valor do observado. Aplicando o modelo NAIVE para tentar fazer a previsão da quantidade de carros vendidos na Noruega, foi-se gerado o seguinte gráfico:

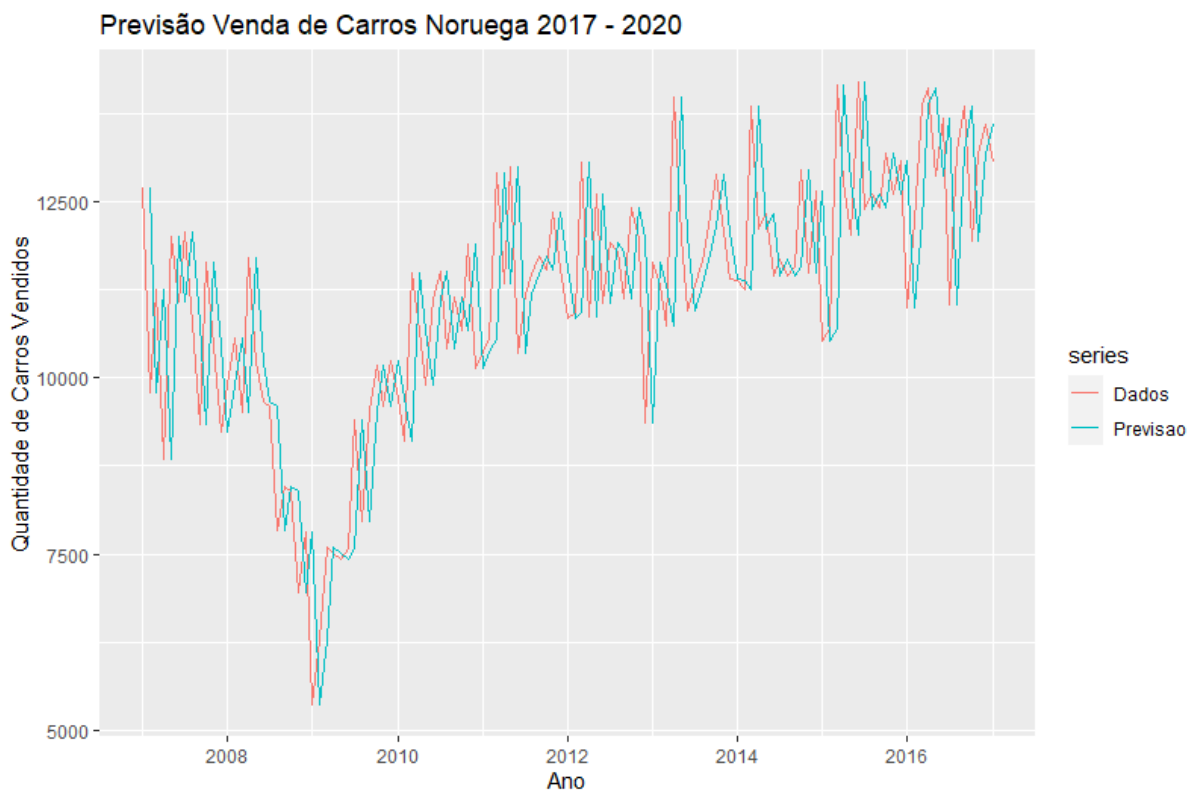


Figura 7.6: NAIVE - Previsão

Com o modelo Naive, é evidenciado que ele foi capaz de prever os dados de maneira correta porém com um pequeno atraso, sendo assim, há indícios de ser um modelo de previsão satisfatório, porém, avaliando a acurácia:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	3.083333	1388.727	1124.35	-0.8068136	10.3567	0.831744	-0.5145416

Figura 7.7: Acurácia - NAIVE

7.2 Análise acurácias

Fazendo a avaliação das acurácias, é possível perceber que os modelos tiveram diversas diferenças entre si. O modelo NAIVE por exemplo, foi o que possui menor *Erro Médio (ME)* porém em relação às outras características foi um dos modelos que obtiveram os maiores números em relação aos outros. Sendo assim, o melhor modelo a ser escolhido é o que possuiu menores números nas características presentes da acurácia, portanto o modelo Holt-Winters com o Método Aditivo ou o modelo ARIMA. Tais modelos foram os que mais bem se comportaram e tiveram os melhores resultados.¹

Dessa maneira, é possível saber a previsão de vendas em unidades de carros para cada mês em cada ano, sendo da seguinte forma:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2017		12785.51	14351.09	14107.95	13806.03	13523.32	13646.08	13310.50	13468.76	14154.93	13638.40	13193.96
2018	12667.44	12915.03	14480.61	14237.47	13935.55	13652.84	13775.60	13440.02	13598.28	14284.45	13767.92	13323.48
2019	12796.96	13044.55	14610.13	14366.99	14065.07	13782.36	13905.12	13569.54	13727.80	14413.97	13897.44	13453.00

Figura 7.8: Holt Winters (Método Aditivo) - Unidades vendidas previsão

	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Feb 2017		12833.08	11476.10	14190.06	10757.753	14908.40
Mar 2017		14212.83	12843.56	15582.10	12118.708	16306.95
Apr 2017		13973.94	12498.22	15449.66	11717.027	16230.86
May 2017		13667.27	12092.29	15242.26	11258.539	16076.00
Jun 2017		13816.71	12148.36	15485.07	11265.182	16368.25
Jul 2017		13254.34	11497.57	15011.11	10567.587	15941.09
Aug 2017		13607.74	11766.79	15448.68	10792.249	16423.22
Sep 2017		13754.27	11832.83	15675.71	10815.683	16692.85
Oct 2017		13735.14	11736.46	15733.83	10678.413	16791.88
Nov 2017		13707.05	11633.98	15780.11	10536.568	16877.52
Dec 2017		13734.34	11589.48	15879.19	10454.066	17014.61
Jan 2018		13249.74	11035.40	15464.07	9863.201	16636.27
Feb 2018		13351.66	10997.99	15705.32	9752.042	16951.27
Mar 2018		14410.01	11985.05	16834.96	10701.359	18118.65
Apr 2018		14200.97	11690.25	16711.70	10361.150	18040.80
May 2018		13932.63	11338.97	16526.29	9965.965	17899.30
Jun 2018		14063.40	11389.37	16737.42	9973.821	18152.97
Jul 2018		13571.31	10819.26	16323.35	9362.410	17780.20
Aug 2018		13880.53	11052.62	16708.45	9555.608	18205.46
Sep 2018		14008.75	11106.95	16910.56	9570.831	18446.68
Oct 2018		13992.02	11018.17	16965.87	9443.905	18540.14
Nov 2018		13967.43	10923.24	17011.63	9311.734	18623.13
Dec 2018		13991.31	10878.37	17104.26	9230.472	18752.15
Jan 2019		13567.28	10387.05	16747.51	8703.534	18431.02

Figura 7.9: ARIMA - Unidades vendidas previsão

¹A data de Janeiro de 2017 não foi incluída pois essa foi a ultima data presente na base de dados.

Analisando os dois modelos dos quais possuíram melhor desempenho (Holt Winters com método aditivo e ARIMA) percebe-se que suas previsões são parecidas para cada mês de determinado ano, entretanto, percebemos que o modelo ARIMA, usando a biblioteca *forecast*, só conseguiu prever até a data de janeiro de 2019, já o modelo Holt Winters com método aditivo previu até o fim do ano de 2019. Porém, o modelo ARIMA conseguiu prever com diversos tipos de possibilidades, com 80% ou 90% de chance para cima (Hi) ou para baixo (Lo), o que não foi encontrado no modelo Holt Winters com Método Aditivo.

Entretanto, como explicado no conceito de Método Aditivo, esse é adequado quando os ciclos de sazonalidade não estão correlacionados ao tempo. Refletindo sobre, é perceptível que a venda de carros está correlacionada com o ano, como é referenciado na figura 3.10 é possível perceber que as variáveis *Quantity* e *Year* possuem correlação positiva, porém em relação ao mês (*Month*) possui correlação neutra. Sendo assim, o modelo Holt-Winters com Método aditivo pode estar equivocado. Dessa maneira, o modelo ARIMA se torna a melhor opção encontrada para a previsão de unidades de carros vendidos na Noruega. 8

Capítulo 8

Conclusão

Dessa maneira é possível concluir, após as análises feitas, que a Noruega possui um grande mercado de vendas de automóveis, e, como visto, a venda de carros tradicionais ao passar dos anos apenas caiu, sendo então, um mercado automotivo se moldando cada vez mais para apenas veículos híbridos e ou elétricos, abandonando os veículos tradicionais, o que pode fazer com que as empresas fabricantes de veículos mudem suas estratégias em relação àquele território.

Também foi possível verificar que a grande economia norueguesa auxiliou para que o país enfrentasse a crise econômica de 2008/2009 no setor automobilístico e em 2010 essa parte da economia do país já voltasse a crescer novamente com números expressivos.

Pode-se concluir que, caso não haja nenhuma outra crise, global ou interna, a Noruega continuará tendo um fértil mercado de automóveis, porém automóveis que sigam a tendência que foi analisada, ou seja, carros elétricos. Marcas e Importações de veículos que não seguirem essa tendência, poderão enfrentar fracasso em relação às vendas de seus veículos.

Referências Bibliográficas

Crawley(2012) Michael J Crawley. *The R book*. John Wiley & Sons. Citado na pág.

Hyndman e Athanasopoulos(2018) Rob J Hyndman e George Athanasopoulos. *Forecasting: principles and practice*. OTexts. Citado na pág.

Morettin e BUSSAB(2017) Pedro Alberto Morettin e WILTON OLIVEIRA BUSSAB. *Estatística básica*. Saraiva Educação SA. Citado na pág.

Provost e Fawcett(2013) Foster Provost e Tom Fawcett. *Data Science for Business: What you need to know about data mining and data-analytic thinking*. "O'Reilly Media, Inc.". Citado na pág.