Relatório Análise de Dados de Viagens

2023-12-07

Instalando as bibliotecas

```
options(repos = c(CRAN = "https://cran.rstudio.com/"))
install.packages('dplyr')
```

```
## Installing package into 'C:/Users/Matheus/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'dplyr' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'dplyr'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Matheus\AppData\Local\R\win-library\4.3\00LOCK\dplyr\libs\x64\dplyr.dll
## to C:\Users\Matheus\AppData\Local\R\win-library\4.3\dplyr\libs\x64\dplyr.dll:
## Permission denied
```

```
## Warning: restored 'dplyr'
```

```
##
## The downloaded binary packages are in
##  C:\Users\Matheus\AppData\Local\Temp\RtmpWA8xhy\downloaded_packages
```

```
install.packages('ggplot2')
```

```
## Installing package into 'C:/Users/Matheus/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Matheus\AppData\Local\Temp\RtmpWA8xhy\downloaded_packages
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Importando os dados

```
dados <- read.csv(
  file = "C:/Users/Matheus/Desktop/MATHEUS/DATA/2019_Viagem.csv",
  sep = ';',
  dec = ',',
  fileEncoding = "latin1"
)
dim(dados)
```

```
## [1] 756704     16
```

O dataframe possui 756.704 linhas e 16 colunas.

Visualização dos dados

```
head(dados,10)
```

```
##    Identificador.do.processo.de.viagem      Situação Código.do.órgão.superior
## 1                             15045825     Realizada                     26000
## 2                             15100682     Realizada                     26000
## 3                             15114708     Realizada                     26000
## 4                             15163874     Realizada                     26000
## 5                             15166192     Realizada                     26000
## 6                             15188479     Realizada                     26000
## 7                             15214826     Realizada                     26000
## 8                             15233002 Não realizada                     52000
## 9                             15238063     Realizada                     52000
## 10                            15238076     Realizada                     52000
##    Nome.do.órgão.superior Código.órgão.solicitante
## 1  Ministério da Educação                    26291
## 2  Ministério da Educação                    26291
## 3  Ministério da Educação                    26291
## 4  Ministério da Educação                    26291
## 5  Ministério da Educação                    26291
## 6  Ministério da Educação                    26291
## 7  Ministério da Educação                    26291
## 8    Ministério da Defesa                    52121
```

```
## 9    Ministério da Defesa                    52121
## 10   Ministério da Defesa                    52121
##                                                 Nome.órgão.solicitante
## 1  Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 2  Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 3  Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 4  Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 5  Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 6  Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 7  Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 8                                                    Comando do Exército
## 9                                                    Comando do Exército
## 10                                                   Comando do Exército
##      CPF.viajante                                Nome
## 1  ***.377.624-** MARINA FERREIRA KITAZONO ANTUNES
## 2  ***.831.975-** JORGE ANDRE DE CARVALHO MENDONCA
## 3  ***.325.718-**       MARCO ANTONIO COUTO JUNIOR
## 4  ***.003.005-**             OLIVAL FREIRE JUNIOR
## 5  ***.660.311-**    CARINA MENDES DOS SANTOS MELO
## 6  ***.655.130-**            RAFAEL RAMIRES JAQUES
## 7  ***.000.460-**     CLARICE MADALENA BUENO ROLIM
## 8  ***.218.347-**  MARCOS ANTONIO AMARO DOS SANTOS
## 9  ***.218.347-**  MARCOS ANTONIO AMARO DOS SANTOS
## 10 ***.218.347-**  MARCOS ANTONIO AMARO DOS SANTOS
##                                  Cargo Período...Data.de.início
## 1                                                    06/02/2019
## 2                                                    01/02/2019
## 3             PESQUISADOR EM GEOCIENCIA              01/02/2019
## 4       PROFESSOR DO MAGISTERIO SUPERIOR             17/02/2019
## 5                             TECNICO I              20/02/2019
## 6  PROFESSOR ENS BASICO TECN TECNOLOGICO             06/03/2019
## 7       PROFESSOR DO MAGISTERIO SUPERIOR             15/02/2019
## 8                                                    08/01/2019
## 9                                                    14/01/2019
## 10                                                   16/01/2019
##    Período...Data.de.fim                    Destinos
## 1             07/02/2019                   Recife/PE
## 2             02/02/2019                   Recife/PE
## 3             01/02/2019                São Paulo/SP
## 4             18/02/2019                 Salvador/BA
## 5             21/02/2019           Rio de Janeiro/RJ
## 6             07/03/2019             Porto Alegre/RS
## 7             16/02/2019             Porto Alegre/RS
## 8             11/01/2019 Porto Alegre/RS, Curitiba/PR
## 9             15/01/2019           Rio de Janeiro/RJ
## 10            17/01/2019                   Recife/PE
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7                                                    Retorno de bolsista do exterior. I
```

```
## 8  O Exmo Sr Gen Ex MARCOS ANTONIO AMARO DOS SANTOS - Sect Econ Fin presidirá as passagens de chefia
## 9                      O Exmo Sr. Gen Ex MARCO ANTONIO AMARO DOS SANTOS - Secretário de Economia e
## 10                                                 O Exmo Sr. Gen Ex MARCOS ANTONIO AMAROS DOS SA
##    Valor.diárias Valor.passagens Valor.outros.gastos
## 1           0.00         3406.33                   0
## 2           0.00         2925.83                   0
## 3           0.00         2760.02                   0
## 4           0.00         2875.92                   0
## 5           0.00         2420.48                   0
## 6           0.00         1262.50                   0
## 7           0.00         2694.58                   0
## 8           0.00         1236.38                   0
## 9         481.65          746.35                   0
## 10        456.30         1293.41                   0
```

Aqui podemos ver o estado das 10 primeiras linhas de dados, e se precisam ser limpos e transformados.

Verificação do tipo dos dados

```
glimpse(dados)
```

```
## Rows: 756,704
## Columns: 16
## $ Identificador.do.processo.de.viagem <int> 15045825, 15100682, 15114708, 1516~
## $ Situação                            <chr> "Realizada", "Realizada", "Realiza~
## $ Código.do.órgão.superior            <int> 26000, 26000, 26000, 26000, 26000,~
## $ Nome.do.órgão.superior              <chr> "Ministério da Educação", "Ministé~
## $ Código.órgão.solicitante            <int> 26291, 26291, 26291, 26291, 26291,~
## $ Nome.órgão.solicitante              <chr> "Fundação Coordenação de Aperfeiço~
## $ CPF.viajante                        <chr> "***.377.624-**", "***.831.975-**"~
## $ Nome                                <chr> "MARINA FERREIRA KITAZONO ANTUNES"~
## $ Cargo                               <chr> "", "", "PESQUISADOR EM GEOCIENCIA~
## $ Período...Data.de.início            <chr> "06/02/2019", "01/02/2019", "01/02~
## $ Período...Data.de.fim               <chr> "07/02/2019", "02/02/2019", "01/02~
## $ Destinos                            <chr> "Recife/PE", "Recife/PE", "São Pau~
## $ Motivo                              <chr> "Regresso de bolsista CAPES do ext~
## $ Valor.diárias                       <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00~
## $ Valor.passagens                     <dbl> 3406.33, 2925.83, 2760.02, 2875.92~
## $ Valor.outros.gastos                 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

Transformação dos dados

Converteremos os dados da coluna Data de Inicio para o tipo Date, faremos isso utilizando o comando as.Date e os colocaremos em uma nova coluna chamada “data.inicio”.

```
dados$data.inicio <- as.Date(dados$Período...Data.de.início, "%d/%m/%Y")

# Formatando a data de inicio - para utilizar apenas Ano/Mês

dados$data.inicio.formatada <- format(dados$data.inicio, "%Y-%m")
```

```r
# Checando os tipos de dados novamente
glimpse(dados)
```

```
## Rows: 756,704
## Columns: 18
## $ Identificador.do.processo.de.viagem <int> 15045825, 15100682, 15114708, 1516~
## $ Situação                            <chr> "Realizada", "Realizada", "Realiza~
## $ Código.do.órgão.superior            <int> 26000, 26000, 26000, 26000, 26000,~
## $ Nome.do.órgão.superior              <chr> "Ministério da Educação", "Ministé~
## $ Código.órgão.solicitante            <int> 26291, 26291, 26291, 26291, 26291,~
## $ Nome.órgão.solicitante              <chr> "Fundação Coordenação de Aperfeiço~
## $ CPF.viajante                        <chr> "***.377.624-**", "***.831.975-**"~
## $ Nome                                <chr> "MARINA FERREIRA KITAZONO ANTUNES"~
## $ Cargo                               <chr> "", "", "PESQUISADOR EM GEOCIENCIA~
## $ Período...Data.de.início            <chr> "06/02/2019", "01/02/2019", "01/02~
## $ Período...Data.de.fim               <chr> "07/02/2019", "02/02/2019", "01/02~
## $ Destinos                            <chr> "Recife/PE", "Recife/PE", "São Pau~
## $ Motivo                              <chr> "Regresso de bolsista CAPES do ext~
## $ Valor.diárias                       <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00~
## $ Valor.passagens                     <dbl> 3406.33, 2925.83, 2760.02, 2875.92~
## $ Valor.outros.gastos                 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ data.inicio                         <date> 2019-02-06, 2019-02-01, 2019-02-0~
## $ data.inicio.formatada               <chr> "2019-02", "2019-02", "2019-02", "~
```

Exploração dos dados

```r
#Verificar se existem valores não preenchidos nas colunas do dados

colSums(is.na(dados))
```

```
## Identificador.do.processo.de.viagem                           Situação
##                                   0                                  0
##            Código.do.órgão.superior             Nome.do.órgão.superior
##                                   0                                  0
##            Código.órgão.solicitante             Nome.órgão.solicitante
##                                   0                                  0
##                        CPF.viajante                               Nome
##                                   0                                  0
##                               Cargo           Período...Data.de.início
##                                   0                                  0
##               Período...Data.de.fim                            Destinos
##                                   0                                  0
##                              Motivo                       Valor.diárias
##                                   0                                  0
##                     Valor.passagens                Valor.outros.gastos
##                                   0                                  0
##                         data.inicio              data.inicio.formatada
##                                   0                                  0
```

```r
# Verificando os valores min, max, e média da coluna sobre valores de passagens

summary(dados$Valor.passagens)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##      0.0      0.0      0.0    590.1    907.0 155531.4
```

Após executar a função summary(), é possivel ver que o valor máximo está muito distante da média, o que indica a presença de outliers. São os valores que fogem da normalidade dos dados, fazendo com que o resultado da análise não mostre a realidade.

```r
# Calculando o desvio padrão

sd(dados$Valor.passagens)
```

```
## [1] 1278.517
```

Continuando a exploração, podemos verificar a quantidade de registros da coluna Situação em cada uma de suas categorias (Realizada, Não realizada). Também podemos obter o percentual das ocorrencias.

```r
# Verificar quantidade de registros em cada categoria

table(dados$Situação)
```

```
##
## Não realizada     Realizada
##         17355        739349
```

```r
# Obtendo os valores em percentual de cada categoria

resultados <- prop.table(table(dados$Situação))*100
resultados_formatados <- sprintf("%.2f%%", resultados)
print(resultados_formatados)
```

```
## [1] "2.29%"  "97.71%"
```

Visualização dos resultados

A visualização dos resultados é a etapa final do projeto acerca da análise dos dados de viagens a serviço. Assim, é o momento de responder às questões:

1 - Qual é o valor gasto por órgão? 2 - Qual é o valor gasto por cidade? 3 - Qual é a quantidade de viagens por mês?

```r
# 1 - Qual é o valor gasto por órgão em passagens?

# Criando um dataframe com os 15 órgãos que gastam mais

p1 <- dados %>%
  group_by(Nome.do.órgão.superior) %>%
  summarise(n = sum(Valor.passagens)) %>%
  arrange(desc(n)) %>%
  top_n(15)
```

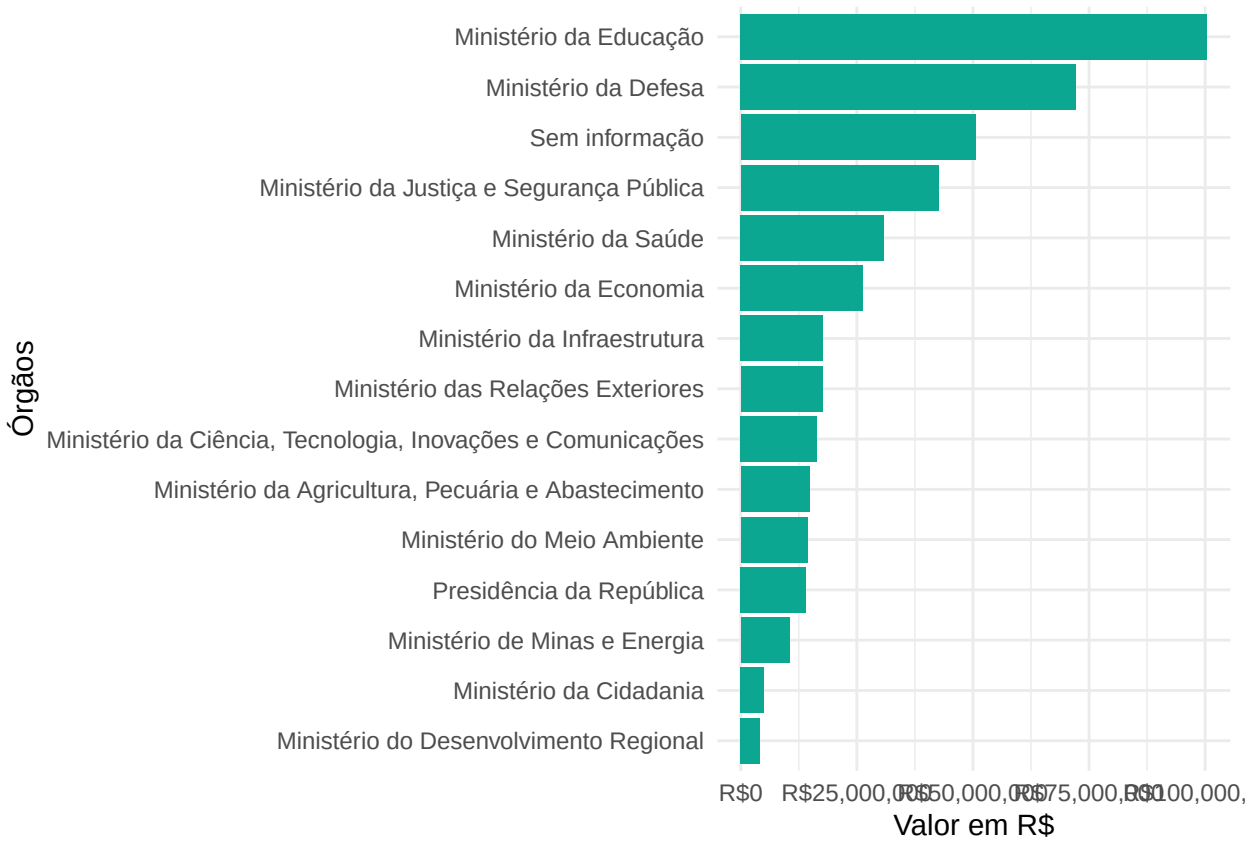
```
## Selecting by n
```

```
# Alterando o nome das colunas
names(p1) <- c("orgao", "valor")

p1
```

```
## # A tibble: 15 x 2
##    orgao                                                    valor
##    <chr>                                                    <dbl>
##  1 Ministério da Educação                                100414641.
##  2 Ministério da Defesa                                   72166086.
##  3 Sem informação                                         50510733.
##  4 Ministério da Justiça e Segurança Pública              42574181.
##  5 Ministério da Saúde                                    30867191.
##  6 Ministério da Economia                                 26361587.
##  7 Ministério da Infraestrutura                           17746202.
##  8 Ministério das Relações Exteriores                     17630741.
##  9 Ministério da Ciência, Tecnologia, Inovações e Comunicações  16428722.
## 10 Ministério da Agricultura, Pecuária e Abastecimento    14840819.
## 11 Ministério do Meio Ambiente                            14419791.
## 12 Presidência da República                               14037021.
## 13 Ministério de Minas e Energia                          10492274.
## 14 Ministério da Cidadania                                 4991945.
## 15 Ministério do Desenvolvimento Regional                  4104182.
```

Aqui, podemos ver quais são os 15 orgãos que tiveram mais gastos com passagens, e o valor gasto por cada um.

```
# Plotando os dados com o ggplot

ggplot(p1, aes(x = reorder(orgao, valor), y = valor)) +
  geom_bar(stat = "identity", fill = "#0ba791") +
  coord_flip() +
  scale_y_continuous(labels = scales::dollar_format(prefix = "R$")) +
  labs(x = "Órgãos", y = "Valor em R$") +
  theme_minimal()
```




```
# 2 - Qual é o valor gasto por cidade?

#Criando um dataframe com as 15 cidades que gastam mais
p2 <- dados %>%
  group_by(Destinos) %>%
  summarise(n = sum(Valor.passagens)) %>%
  arrange(desc(n)) %>%
  top_n(15)
```

```
## Selecting by n
```

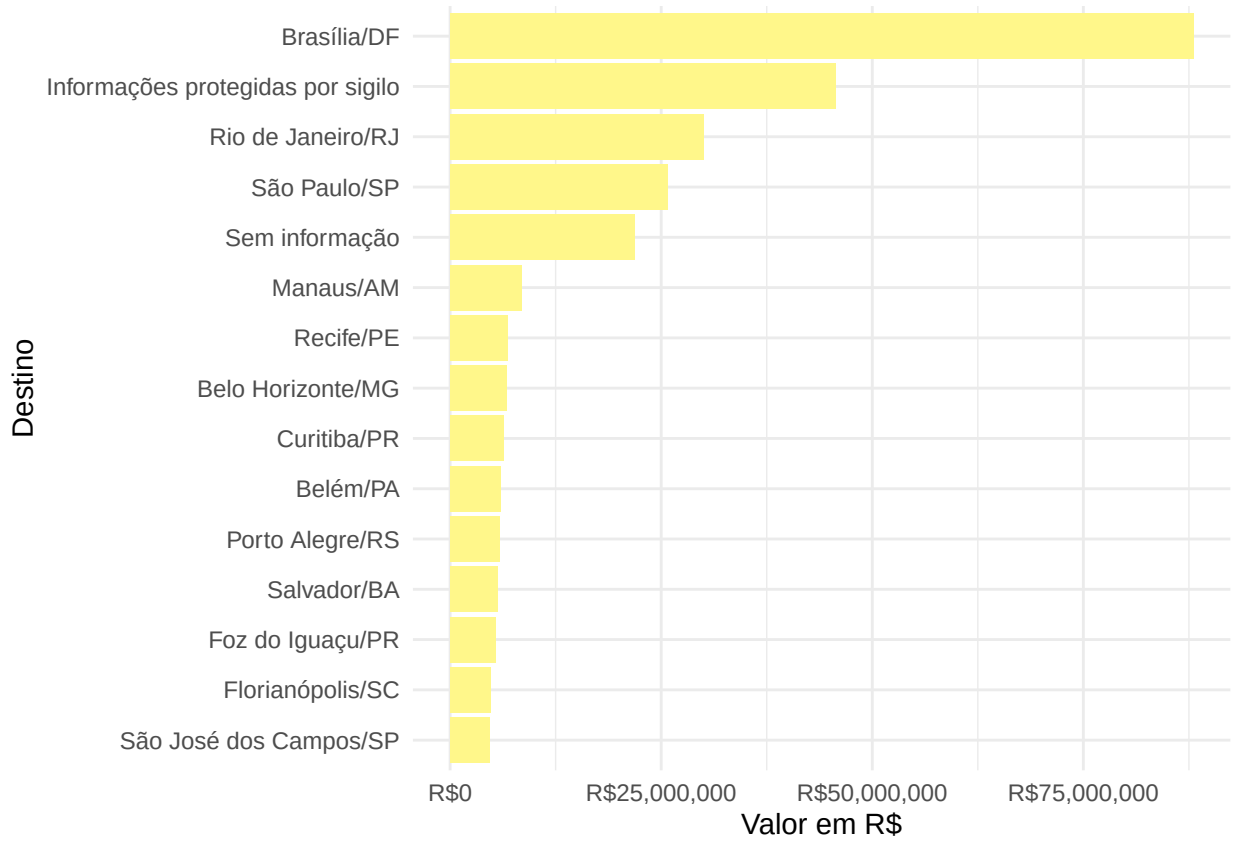
```
#Alterando o nome das colunas
names(p2) <- c("destino", "valor")

p2
```

```
## # A tibble: 15 x 2
##    destino                              valor
##    <chr>                                <dbl>
##  1 Brasília/DF                        88009638.
##  2 Informações protegidas por sigilo  45675238.
##  3 Rio de Janeiro/RJ                  29999886.
##  4 São Paulo/SP                       25764054.
##  5 Sem informação                     21891215.
##  6 Manaus/AM                           8463212.
```

```
##  7 Recife/PE                        6833241.
##  8 Belo Horizonte/MG                6691385.
##  9 Curitiba/PR                      6388793.
## 10 Belém/PA                         5982406.
## 11 Porto Alegre/RS                  5921046.
## 12 Salvador/BA                      5634670.
## 13 Foz do Iguaçu/PR                 5359965.
## 14 Florianópolis/SC                 4746875.
## 15 São José dos Campos/SP           4709705.
```

Aqui podemos ver quais foram as 15 cidades que mais gastaram, e o valor gasto por cada uma.

```r
#Criando o gráfico
ggplot(p2, aes(x = reorder(destino, valor), y = valor))+
  geom_bar(stat = "identity", fill = "#FFF78A")+
  coord_flip()+
  scale_y_continuous(labels = scales::dollar_format(prefix = "R$")) +
  labs(x = "Destino", y = "Valor em R$")+
  theme_minimal()
```



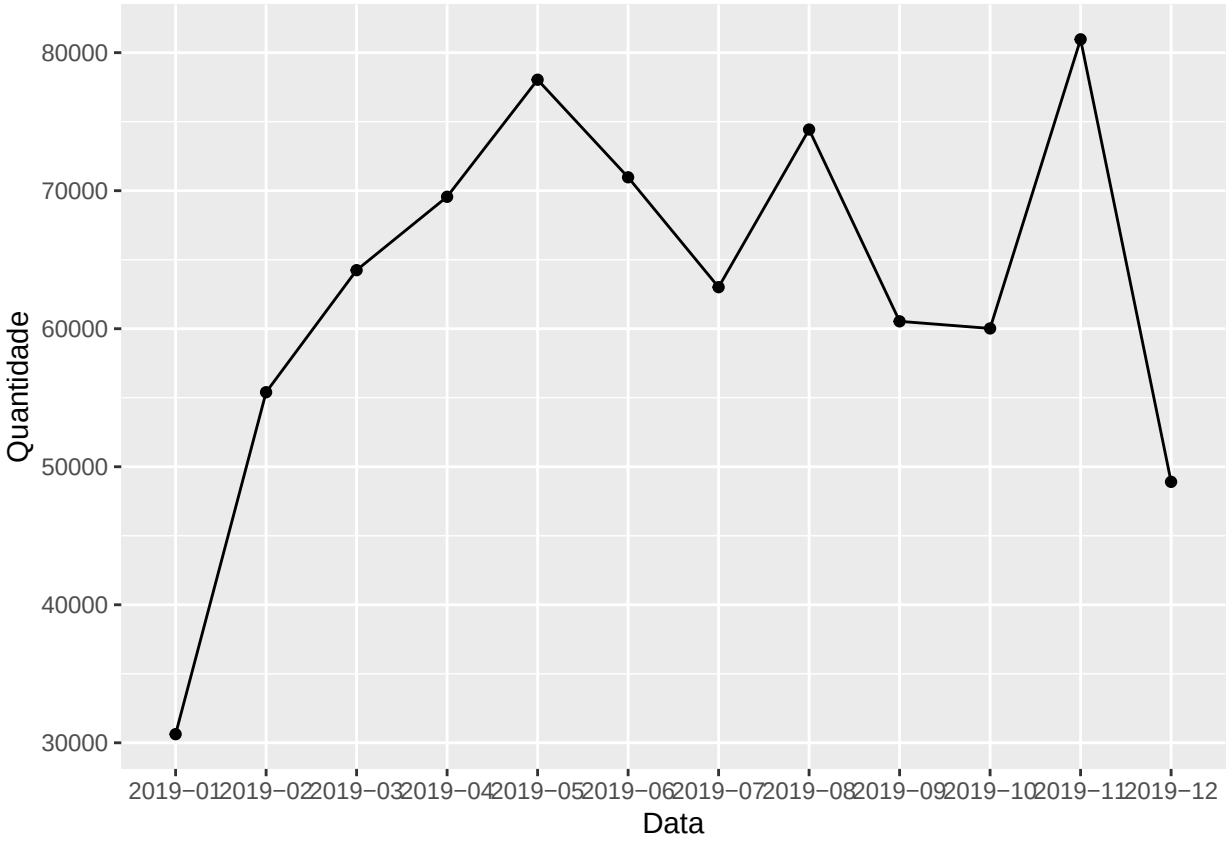

```r
# 3 - Qual é a quantidade de viagens por mês?

#Criando um dataframe com a quantidade de viagens por Ano/mês
p3 <- dados %>%
  group_by(data.inicio.formatada) %>%
```

```r
  summarise(qtd = n_distinct(Identificador.do.processo.de.viagem))

head(p3,12)
```

```
## # A tibble: 12 x 2
##    data.inicio.formatada   qtd
##    <chr>                 <int>
##  1 2019-01               30626
##  2 2019-02               55399
##  3 2019-03               64242
##  4 2019-04               69559
##  5 2019-05               78039
##  6 2019-06               70971
##  7 2019-07               63009
##  8 2019-08               74432
##  9 2019-09               60535
## 10 2019-10               60021
## 11 2019-11               80965
## 12 2019-12               48906
```

Aqui podemos ver a quantidade total de viagens realizadas por mês.

```r
#Criando o gráfico
ggplot(p3, aes(x = data.inicio.formatada, y = qtd, group = 1)) +
  geom_line() +
  geom_point() +
  labs(x = "Data", y = "Quantidade")
```




```
install.packages("rmarkdown")
```

```
## Installing package into 'C:/Users/Matheus/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'rmarkdown' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Matheus\AppData\Local\Temp\RtmpWA8xhy\downloaded_packages
```

```
install.packages('tinytex')
```

```
## Installing package into 'C:/Users/Matheus/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'tinytex' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Matheus\AppData\Local\Temp\RtmpWA8xhy\downloaded_packages
```

```
library(tinytex)
```