RELATÓRIO DE ANÁLISE DESCRITIVA

FORTALEZA
2018

1 Análise das despesas

```
##setwd("projetos/Consulta Mobills/")
Despesas <- fread("./data/dadosReceitasUsers/Despesas20190711.csv")
Despesas2018 <- fread("./data/dadosReceitasUsers/Despesas20180107.csv")
Despesas <- do.call("rbind", list(Despesas,Despesas2018))
Despesas %>% mutate(ano= lubridate::year(DataDespesa)) -> Despesas
```

As medidas de resumo mostram que claramente há valores inválidos e outliers nos dados.

```
Despesas$Valor <- gsub(pattern = ",",replacement = ".", Despesas$Valor)
Despesas$Valor <- as.numeric(Despesas$Valor)
```

```
Despesas %>%
  select(ano,Valor)%>%
  split(.$ano) %>%
  map(summary)
```

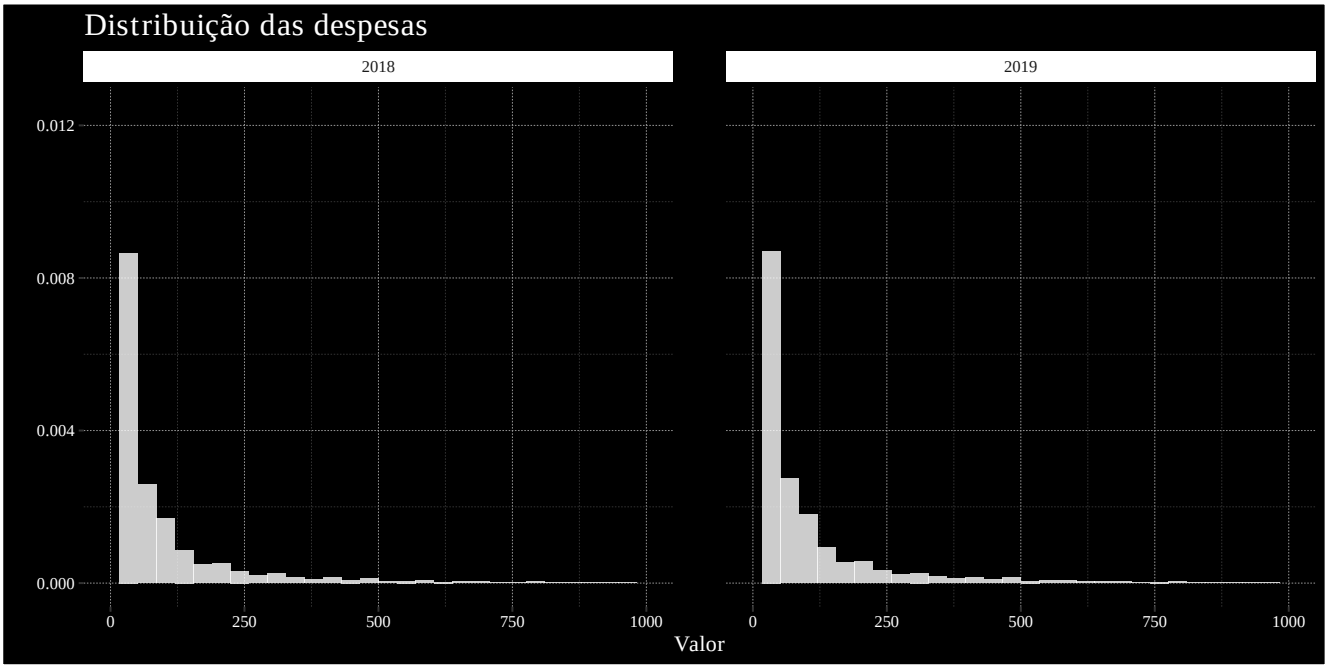
```
## $`2018`
##       ano           Valor
##  Min.   :2018   Min.   : -10000
##  1st Qu.:2018   1st Qu.:      9
##  Median :2018   Median :     22
##  Mean   :2018   Mean   :    133
##  3rd Qu.:2018   3rd Qu.:     60
##  Max.   :2018   Max.   :6000000
##
## $`2019`
##       ano           Valor
##  Min.   :2019   Min.   :          -15591
##  1st Qu.:2019   1st Qu.:              10
##  Median :2019   Median :              25
##  Mean   :2019   Mean   :     12899026578
##  3rd Qu.:2019   3rd Qu.:              69
##  Max.   :2019   Max.   :9000000000000000
```

Para plotar o Histograma dos Valores gastos(Despesas) vamos limitar a variável ‘Valor‘ em até 1000 reais. Tendo em vista que quase a totalidade dos dadados se concentram nesse intervalo.

```
Despesas %>%
  group_by(UsuarioId,
           dia = lubridate::day(DataDespesa),
           mes = lubridate::month(DataDespesa),
           ano = lubridate::year(DataDespesa)) %>%
    summarise(count = n(), valorSoma= sum(Valor)) %>%
    arrange(desc(valorSoma)) -> desp
```

```
Despesas %>% dplyr::filter(Valor > 0 & Valor < 1000) %>%
    ggplot(aes(Valor,y=..density..))+
    geom_histogram(fill="white",alpha=0.8)+
    facet_grid(~ano )+temaMobills+
        labs(title="Distribuição das despesas")+
    scale_x_continuous(limits=c(0,1000))
```

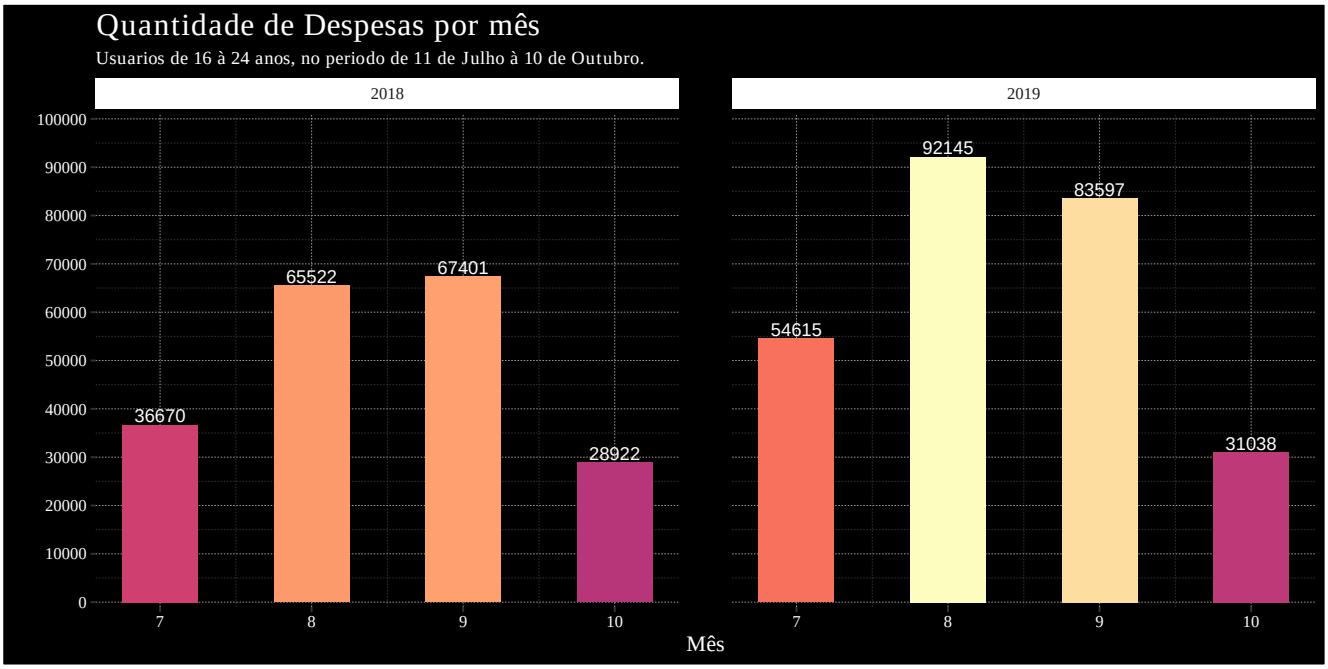
```
## Warning: Removed 4 rows containing missing values (geom_bar).
```



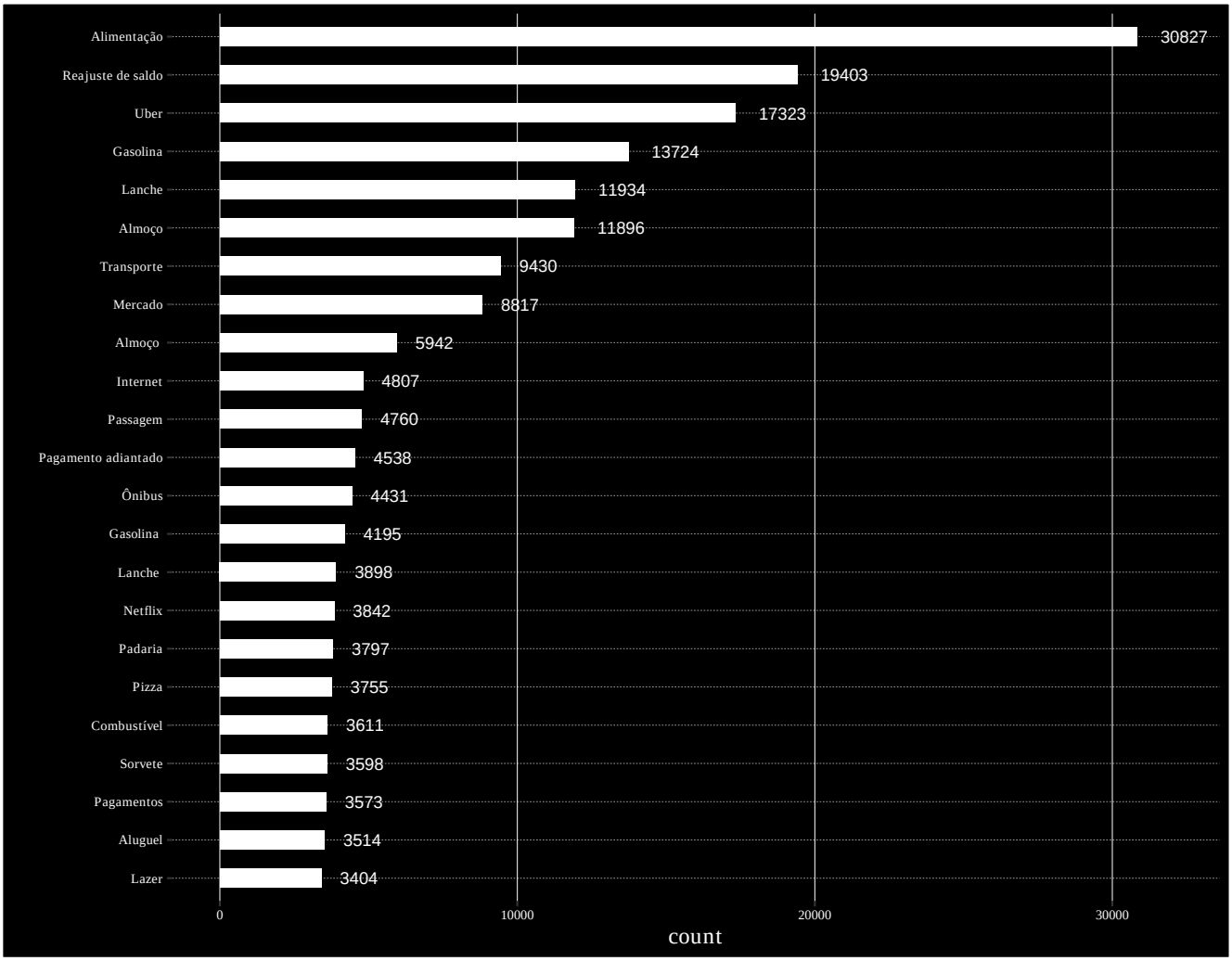

```
##histogram by month
Despesas %>% dplyr::filter(Valor > 0 & Valor < 1000) %>%
    ggplot(aes(Valor,y=..density..))+
    geom_histogram(fill="white",alpha=0.8)+
    facet_grid(~ano )+temaMobills+
        labs(title="Distribuição das despesas")+
    scale_x_continuous(limits=c(0,1000))+
  facet_grid(~mes)


## Error:  At least one layer must contain all faceting variables:  'mes'.
## * Plot is missing 'mes'
## * Layer 1 is missing 'mes'
```

```
desp %>%
  group_by(mes,ano) %>%
  summarise(contagem=n()) %>%
  ggplot(aes(mes, contagem,labs=contagem))+
  geom_col(aes(fill=contagem),
           width = 0.5)+
  scale_fill_viridis(option="magma",begin=0.5)+
  labs(title="Quantidade de Despesas por mês",
```

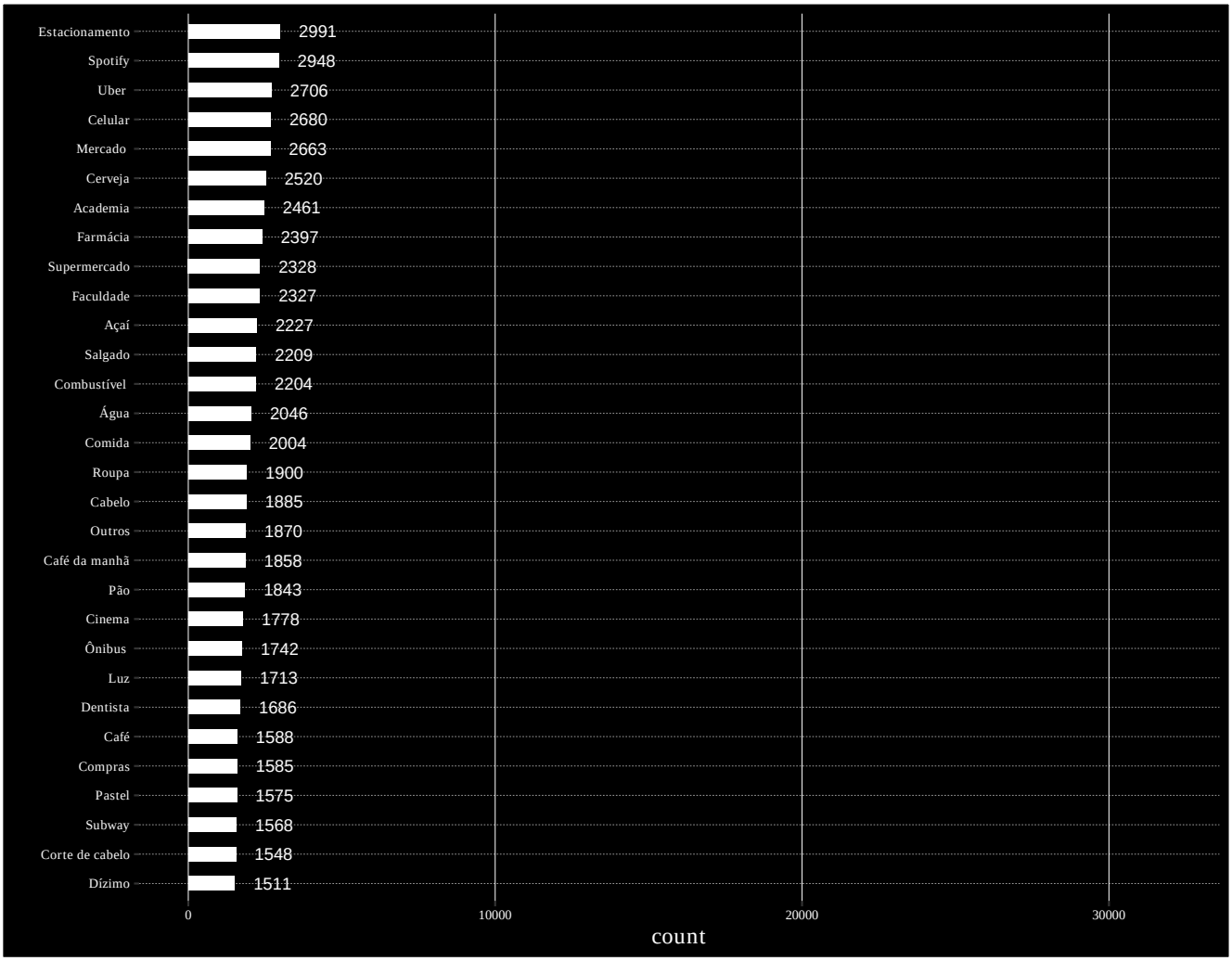
```
      subtitle = "Usuarios de 16 à 24 anos, no periodo de 11 de Julho à 10 de Outubro.",
      x="Mês",
      y="Quantidade de despesas")+
  temaMobills+
  scale_y_continuous(limits = c(0,100000),
                     expand=c(0.01009,0.000000001),
                     breaks = seq(0,150000,10000))+
  geom_text(aes(label=contagem),
            size=3.5,
            colour="white",
            vjust=-0.2)+
  facet_grid(~ano)
```




```
Despesas %>%
    group_by(Descricao) %>%
    summarise(count = n(), valorSoma= sum(Valor)) %>%
    top_n(100000) %>% filter(count > 3000) %>% arrange(desc(count))%>%
  ggplot(aes(x=reorder(Descricao,count,max),count),labels=count)+geom_col(fill="white",width = 0.5)+
  coord_flip()+
  temaMobills+
  theme(axis.text = element_text(size=7),
        panel.grid.major.x =element_line(colour="white",linetype = 1),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(size=0.1))+
  geom_text(aes(label=count),colour="white",size=3,hjust=-0.5)+
  scale_y_continuous(limits=c(0,32000))
```




```
Despesas %>%
    group_by(Descricao) %>%
    summarise(count = n(), valorSoma= sum(Valor)) %>%
    top_n(100000) %>% filter(count <3000,count>1500) %>% arrange(desc(count))%>%
  ggplot(aes(x=reorder(Descricao,count,max),count),labels=count)+geom_col(fill="white",width = 0.5)+
  coord_flip()+
  temaMobills+
  theme(axis.text = element_text(size=7),
        panel.grid.major.x =element_line(colour="white",linetype = 1),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(size=0.1))+
  geom_text(aes(label=count),colour="white",size=3,hjust=-0.5)+
  scale_y_continuous(limits=c(0,32000))
```


Estacionamento 2991
Spotify 2948
Uber 2706
Celular 2680
Mercado 2663
Cerveja 2520
Academia 2461
Farmácia 2397
Supermercado 2328
Faculdade 2327
Açaí 2227
Salgado 2209
Combustível 2204
Água 2046
Comida 2004
Roupa 1900
Cabelo 1885
Outros 1870
Café da manhã 1858
Pão 1843
Cinema 1778
Ônibus 1742
Luz 1713
Dentista 1686
Café 1588
Compras 1585
Pastel 1575
Subway 1568
Corte de cabelo 1548
Dízimo 1511
0
10000
20000
30000
count