

# Hands-on R - Análises descritivas básicas

Rodrigo Heldt - rodrigoheldt@gmail.com

IMED - HOPEAD - Nov 13, 2017

## Análises descritivas básicas

### 0. Importando a base de dados de um arquivo csv

```
base = read.csv("C:/Users/Rodrigo/Dropbox/Hands-on R Workshop/Parte 2 - Pre-processamento dos dados/Dat  
sep = ",")
```

### 1. Análises descritivas

```
## Media  
mean(base$SalarioPorHora)
```

```
## [1] 114.9
```

```
## Desvio padrao
```

```
sd(base$SalarioPorHora)
```

```
## [1] 55.15605
```

```
## Minimo
```

```
min(base$SalarioPorHora)
```

```
## [1] 32
```

```
## Maximo
```

```
max(base$SalarioPorHora)
```

```
## [1] 384
```

```
## Mediana
```

```
median(base$SalarioPorHora)
```

```
## [1] 100
```

```
## Resumo com diversas estatísticas descritivas:
```

```
## Minimo, Primeiro Quartil, Mediana, Mediam, Terceiro Quartil e Maximo
```

```
summary(base$SalarioPorHora)
```

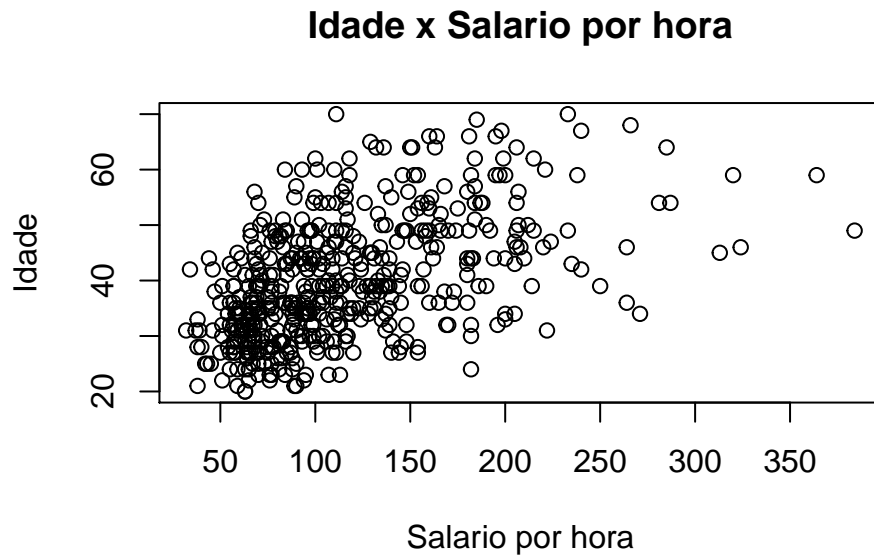
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##      32.0   72.0   100.0   114.9   144.0   384.0
```

### 2. Análises gráficas básicas

```
## Scatter plot de SalarioPorHora e Idade
```

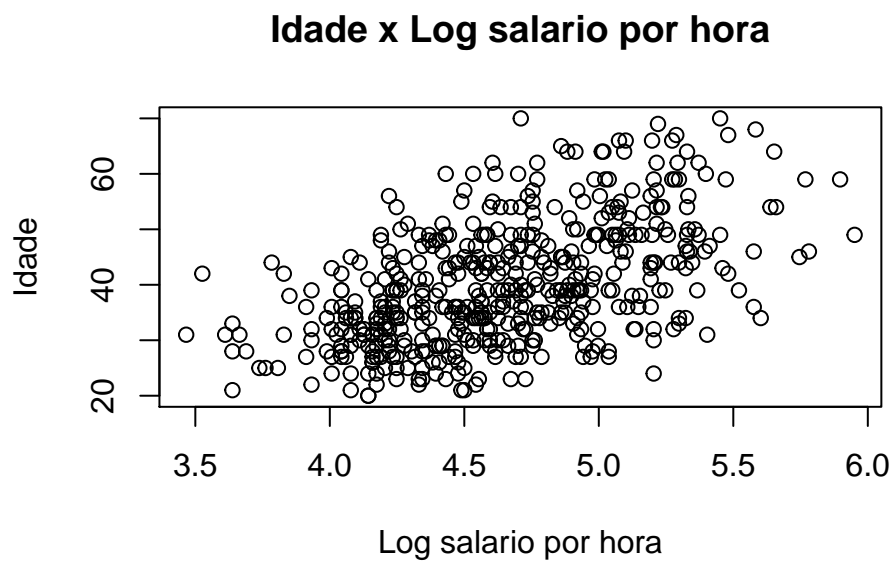
```
plot(base$SalarioPorHora,  
      base$Idade,
```

```
main = "Idade x Salario por hora",  
xlab = "Salario por hora",  
ylab = "Idade")
```

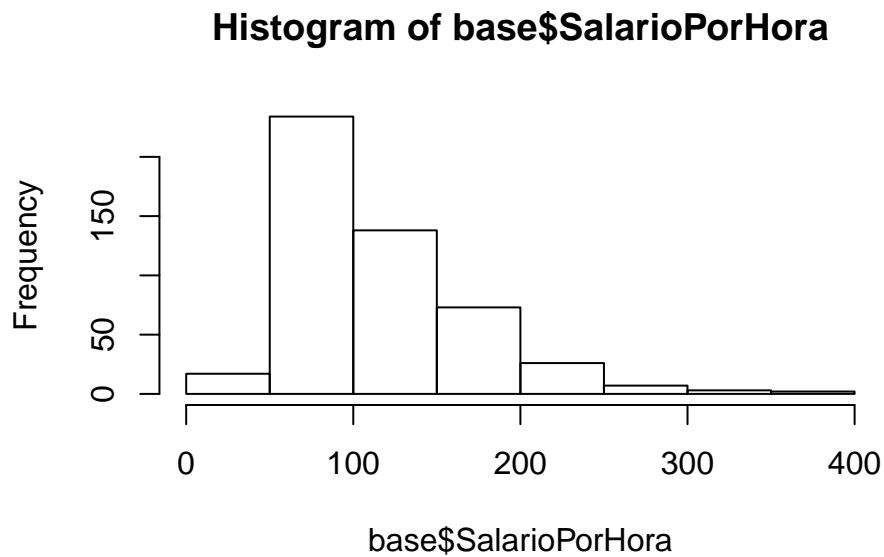


```
## Scatter plot de LogSalarioPorHora e Idade
```

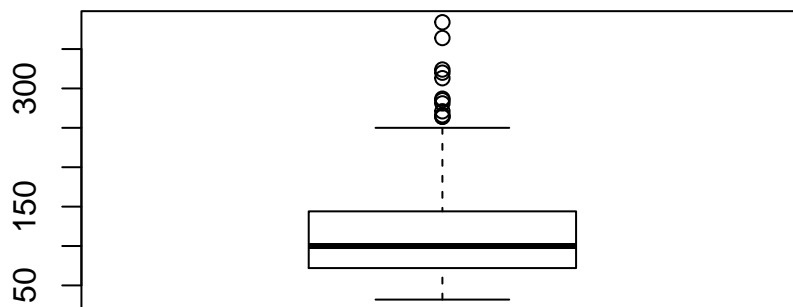
```
plot(base$LogSalarioPorHora,  
base$Idade,  
main = "Idade x Log salario por hora",  
xlab = "Log salario por hora",  
ylab = "Idade")
```



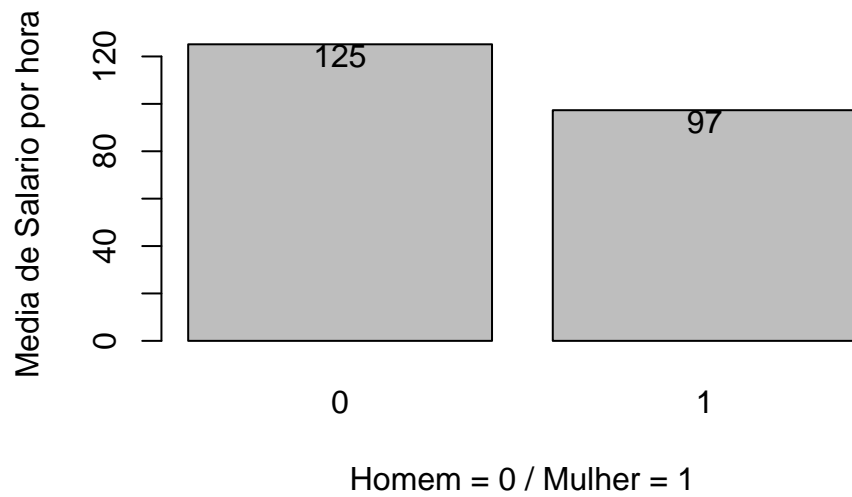
```
## Histograma
hist(base$SalarioPorHora)
```



```
## Boxplot
boxplot(base$SalarioPorHora)
```



```
# Grafico de barras com as medias do SalarioPorHora por sexo
media.salario.sexo <- by(data = base$SalarioPorHora, INDICES = base$Mulher, FUN = mean)
bb <- barplot(media.salario.sexo,
              xlab = "Homem = 0 / Mulher = 1",
              ylab = "Media de Salario por hora")
text(bb, media.salario.sexo-5, labels=round(media.salario.sexo))
```



### 3. Exercício

E1 - Aplique a função `summary` sobre a variável `Idade` para verificar as medidas descritivas

E2 - Calcule o desvio padrão da variável `Idade`.

E3 - Fazer um histograma da variável `LogSalarioPorHora`.