Introdução e Estatística Descritiva

Fernando B. Sabino da Silva

Contents

0.1 Rstudio

- Faça uma pasta no seu computador onde você deseja manter os arquivos para usar no **Rstudio**.
- Defina o diretório de trabalho nesta pasta: `Session -> Set Working Directory -> Choose Directory` (atalho: Ctrl+Shift+H).
- Torne a alteração permanente definindo o diretório padrão em: `Tools -> Global Options -> Choose Directory`.

0.2 R básico

- Cálculos simples:

```
4.6 * (2 + 3)^4
```

```
## [1] 2875
```

- Defina um objeto (escalar) e o imprima:

```
a <- 4
a
```

```
## [1] 4
```

- Defina um objeto (vetor) e o imprima:

```
b <- c(2, 5, 7)
b
```

```
## [1] 2 5 7
```

- Defina uma sequência de números e a imprima:

```
s <- 1:4
s
```

```
## [1] 1 2 3 4
```

- Nota: Um comando mais flexível para sequências:

```
s <- seq(1, 4, by = 1)
```

- **R** faz cálculos elemento a elemento:

```
a * b
```

```
## [1]  8 20 28
```

```
a + b
```

```
## [1]  6  9 11
```

```
b ^ 2
```

```
## [1]  4 25 49
```

- Soma e produto de elementos:

```
sum(b)
```

```
## [1] 14
```

```
prod(b)
```

```
## [1] 70
```

0.3 Extensões do R

- O **R** não precisa ser usado apenas como calculadora ou para atribuição de objetos simples. A sua funcionalidade pode ser estendida atráves de bibliotecas ou pacotes (muito similar a utilização de Plug-ins nos navegadores ou baixar aplicativos no google play). Alguns já vem instalados (automaticamente, by default) no **R** e você precisa apenas carregá-los (como fazemos depois que baixamos um aplicativo no celular e queremos usá-lo, por exemplo).
- Para instalar um novo pacote no **Rstudio** você pode usar o menu: `Tools -> Install Packages`
- Você precisa saber o nome do pacote que deseja instalar. Você também pode fazê-lo através do comando install.packages como abaixo:

```
install.packages("mosaic")
```

- Uma vez que o pacote esteja instalado, você pode carregá-lo através do comando `library` (ou `require`):

```
library(mosaic)
```

```
## Warning: package 'mosaic' was built under R version 3.4.3
```

```
## Warning: package 'dplyr' was built under R version 3.4.3
```

```
## Warning: package 'ggformula' was built under R version 3.4.3
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

```
## Warning: package 'mosaicData' was built under R version 3.4.3
```

- Isto carrega o pacote `mosaic` que possuí muitas funções convenientes para este curso (voltaremos a isso mais tarde). Ele também imprime muitas informações sobre as funções que foram alteradas pelo pacote `mosaic`, mas você pode ignorar isto com segurança.

0.4 Ajuda do R

- Você pode receber ajuda (help) via `?<command>`:

```
?sum
```

- Procurando por ajuda:

```
help.search("plot")
```

- Você pode encontrar um cheat sheet com funções do **R** que usaremos neste curso aqui. Caso o arquivo não apareça, clique com o botão direito em cima do link e escolha `Open link in a new tab`.
- Você pode salvar os comandos que você porventura tenha digitado em um arquivo para uso posterior:
  - Selecione o guia History no painel superior direito no **Rstudio** .
  - Marque os comandos que você deseja salvar.
  - Pressione o botão `To Source`.
- Pratique as suas habilidades básicas em: http://tryr.codeschool.com

0.5 Dados: Exemplos

- Data: Legibilidade de Anúncios em Revistas
- Trinta revistas foram classificadas pelo nível educacional de seus leitores.
- Três revistas foram selecionadas **aleatoriamente** de cada um dos seguintes grupos:
  - Grupo 1: maior nível educacional
  - Grupo 2: nível educacional médio
  - Grupo 3: nível educacional mais baixo.
- Seis anúncios foram selecionados **aleatoriamente** de cada uma das nove revistas selecionadas:

  - Grupo 1: [1] Scientific American, [2] Fortune, [3] The New Yorker
  - Grupo 2: [4] Sports Illustrated, [5] Newsweek, [6] People
  - Grupo 3: [7] National Enquirer, [8] Grit, [9] True Confessions
- Logo, os dados contém informações sobre um total de 54 anúncios.

0.6 Objetivos do Capítulo

- Identificar o tipo de variável (por exemplo, numérica ou categórica; discreta ou contínua; ordenada ou não)
- Usar visualizações apropriadas para diferentes tipos de dados (por exemplo, histograma, gráfico de barras (barplot), gráfico de dispersão (scatterplot), boxplot, etc.)
- Criar e interpretar tabelas de contingência e de distribuições de frequência (tabelas uni e bidirecionais - de uma e duas entradas)
- Usar diferentes medidas de tendência central e dispersão e ser capaz de descrever a robustez de diferentes estatística (por exemplo, quando devemos usar cada uma e até que ponto elas podem ser usadas)
- Descrever a forma das distribuições (usando também gráficos como o histograma e o boxplot)

0.7 Exemplo (continuação) - variáveis e formato

- Para cada anúncio (54 casos), os dados abaixo foram observados.
- **Nome das variáveis**:
  - WDS = número de palavras na propaganda
  - SEN = número de frases na propaganda
  - 3SYL = número de palavras com 3 ou mais sílabas no anúncio
  - MAG = revista (1 a 9 como na página anterior)
  - GROUP = nível educacional (1 a 3 como na página anterior)
- Dê uma olhada nos dados usando **Rstudio**:

```
magAds <- read.delim("C:/Users/fsabino/Desktop/Codes/papers/Introductory_Stat_I/notebook/datasets_ads.t
head(magAds)
```

```
##   WDS SEN X3SYL MAG GROUP
## 1 205   9    34   1     1
## 2 203  20    21   1     1
## 3 229  18    37   1     1
## 4 208  16    31   1     1
## 5 146   9    10   1     1
## 6 230  16    24   1     1
```

- Os nomes das variáveis estão na linha superior. Não é permitido começar o nome de uma variável com um dígito, então um `X` foi adicionado em `X3SYL`.

0.8 Tipos de Dados

0.8.1 Variáveis Quantitativas

- Medições contém valores numéricos.
- Os dados quantitativos geralmente surgem das seguintes maneiras:
  - **Variáveis contínuas**: `medições` de, por exemplo, tempo de espera em uma fila, receitas, preços de ações, etc.
  - **Variáveis discretas**: `contagens` de, por exemplo, palavras em um texto, acessos de um website, números de chegadas em uma fila em uma hora, etc.
- Medidas como esta têm um escala bem definida e no **R** elas são armazenadas como numéricas (**numeric**).

0.8.2 Variáveis Categóricas/Qualitativas

- A medida é um fator proveniente de um conjunto de determinadas categorias. Exemplos: sexo (masculino/feminino), classe social, escore de satisfação (baixo/médio/alto), etc.
- A medida é normalmente armazenada (o que é altamente recomendável) como um fator (**factor**) no **R**. As categorias possíveis são chamadas de níveis (**levels**). Examplo: os níveis do fator "sexo" são masculino/feminino.
- Fatores têm duas possíveis escalas:
  - **Escala Nominal**: Não há ordenação natural entre os níveis dos fatores. Exemplos: sexo e cor do cabelo.
  - **Escala Ordinal**: Há uma ordenação natural entre os níveis dos fatores. Exemplos: classe social e escore de satisfação. Um fator no **R** pode ter um chamado atributo (**attribute**) atribuído, informando que a escala é ordinal (veja a função ordered()).

1 População e Amostra

1.1 Objetivo da Estatística

- O objetivo da Estatística é "dizer algo" sobre a população.
- Tipicamente, isso é feito utilizando as informações de uma amostra aleatória retirada da população de interesse.
- Antes de retirar a amostra podemos ter alguma hipótese sobre a população. A amostra é então analisada como o objetivo de testar esta hipótese.
- O processo de fazer conclusões para uma população com base em uma amostra é chamado de **inferência estatística**.

1.2 Seleção aleatória

- Exemplo: Para os dados das revistas:
  - Primeiro nós selecionamos **aleatoriamente** 3 revistas de cada grupo.
  - Na sequência, nós selecionamos, **aleatoriamente**, 6 anúncios de cada revista.
  - Um detalhe importante é que a seleção é feita de maneira completamente **aleatória**, i.e.
    - cada revista dentro de um grupo tem a mesma chance de ser escolhida e
    - cada anúncio dentro de uma revista tem a mesma chance de ser escolhido.
- No que veremos neste curso é fundamental que os dados coletados respeitem o princípio da aleatoriedade. Sempre que utilizarmos a palavra **amostra** daqui em diante, estaremos nos referindo a uma a.a. (amostra aleatória).
- Mais geralmente:
  - Nós temos uma **população** de objetos.
  - Nós escolhemos aleatoriamente n destes objetos, e do j-ésimo objeto nós obtemos a medição y_j, $j = 1, 2, \ldots, n$.
  - As medições $y_1, y_2, \ldots, y_n$ são então chamadas de **amostra**. Só uma amostra (que contém n elementos) e não várias amostras.
- Se nós, por exemplo, estivermos medindo a qualidade da água 4 vezes em um ano é uma má ideia coletarmos dados apenas com tempo bom. A amostragem escolhida ao longo do tempo não pode ser influenciada por algo que possa influenciar a medida em si.

2 Tabelas de agrupamento e frequência

2.1 Dividir toda a gama de valores em uma série de intervalos: "Binning"

- A função `cut` irá dividir o intervalo de uma variável numérica em vários intervalos de tamanho igual e registrar a qual intervalo pertence cada observação. Por exemplo, para a variável `X3SYL` (o número de palavras com mais de 3 sílabas):

```
# Antes de 'cortar':
magAds$X3SYL[1:5]
```

```
## [1] 34 21 37 31 10
```

```
# Após 'cortar' (dividir) em 4 intervalos:
syll <- cut(magAds$X3SYL, 4)
syll[1:5]
```

```
## [1] (32.2,43]     (10.8,21.5]   (32.2,43]     (21.5,32.2]   (-0.043,10.8]
## Levels: (-0.043,10.8] (10.8,21.5] (21.5,32.2] (32.2,43]
```

- O resultado é um fator (`factor`) e os rótulos são os intervalos. Os itens personalizados podem ser atribuídos através do argumento `labels` (rótulos):

```
labs <- c("poucas", "algumas", "muitas", "demais")
syll <- cut(magAds$X3SYL, 4, labels = labs) # Nota: isso sobreescreverá a variável 'syll' definida acim
syll[1:5]
```

```
## [1] demais  algumas demais  muitas  poucas
## Levels: poucas algumas muitas demais
```

```
magAds$syll <- syll # Adicionando uma nova coluna ao conjunto de dados
```

2.2 Tabelas

- Para resumir os resultados nós podemos utilizar a função `tally` (contagem) do pacote`mosaic` (relembre que o pacote **deve ser carregado** escrevendo `library(mosaic)` se você ainda não o fez):

```
tally( ~ syll, data = magAds)
```

```
## syll
##  poucas algumas  muitas  demais
##      26      14      10       4
```

- Em porcentagem:

```
tally( ~ syll, data = magAds, format = "percent")
```

```
## syll
##  poucas algumas  muitas  demais
##    48.1    25.9    18.5     7.4
```

- Aqui nós usamos uma `fórmula` (caracterizada pelo til) para indicar que nós queremos a variável syll do conjunto de dados `magAds` (sem o til o **R** iria procurar por uma variável global chamada `syll` caso ela exista (se não existir dará certo) e a utilizaria ao invés da que queremos).

2.3 2 fatores: Tabulação Cruzada

- Para fazer uma tabela da combinação de dois fatores nós utilizamos a função `tally` novamente:

```
tally( ~ syll + GROUP, data = magAds)
```

```
##          GROUP
## syll       1  2  3
##   poucas   8 11  7
##   algumas  4  2  8
##   muitas   3  5  2
##   demais   3  0  1
```

- Frequências relativas (em porcentagem) por coluna:

```
tally( ~ syll | GROUP, data = magAds, format = "percent")
```

```
##          GROUP
## syll         1    2    3
##   poucas  44.4 61.1 38.9
##   algumas 22.2 11.1 44.4
##   muitas  16.7 27.8 11.1
##   demais  16.7  0.0  5.6
```

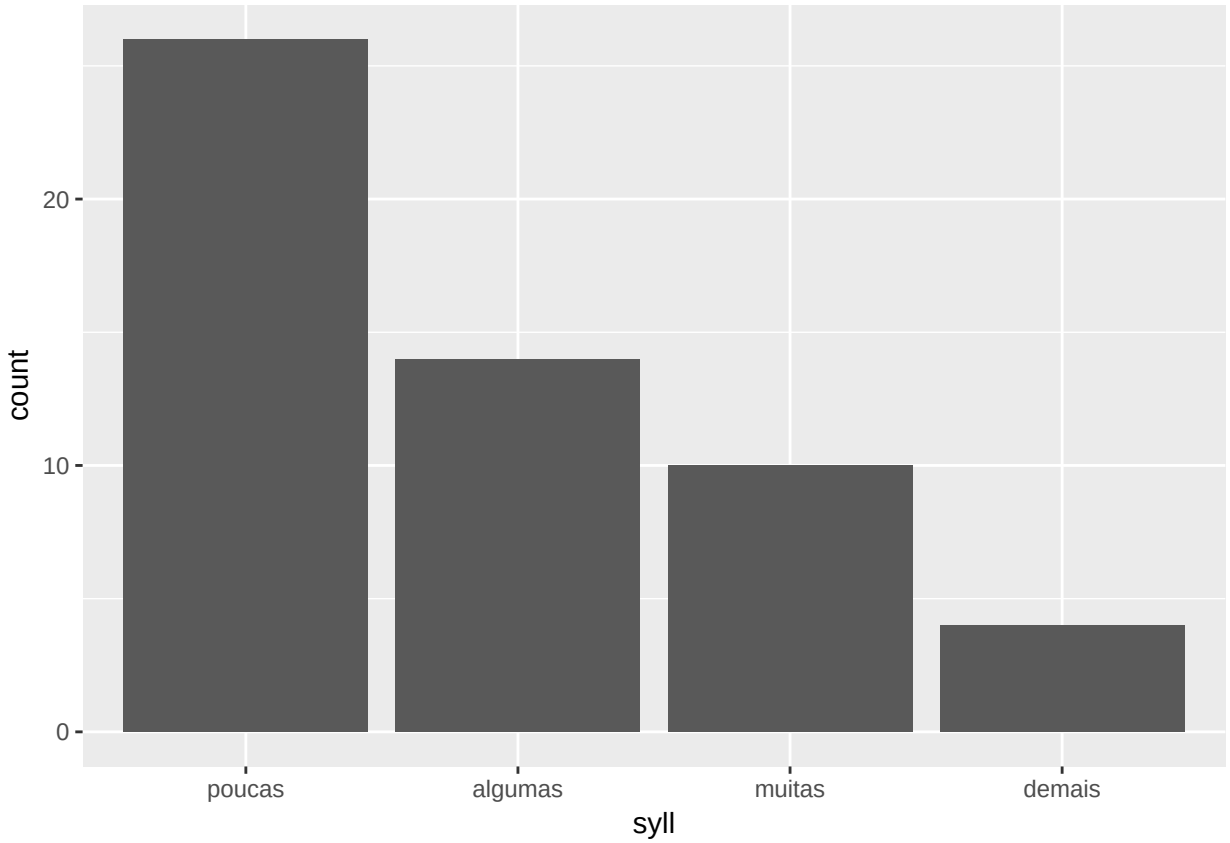
- A tabela acima mostra, por exemplo, qual a porcentagem de anúncios no grupo 1 que tem 'poucas', 'algumas', 'muitas' ou 'demais' com mais de 3 sílabas.

3 Gráficos

3.1 Gráfico de barras

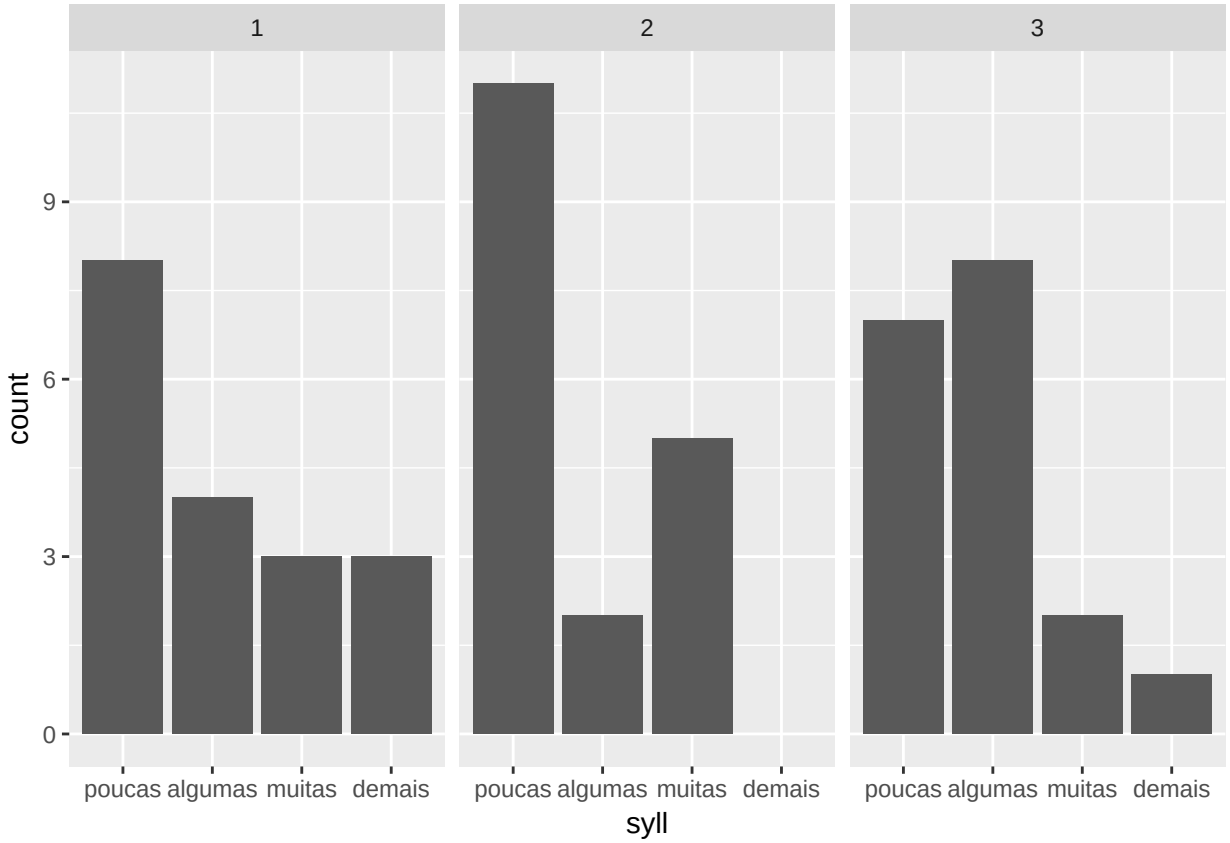
- Para criar um gráfico de barras com os dados da tabela nós usamos a função `gf_bar` do pacote `mosaic`. Para cada nível do fator uma caixa é desenhada com a altura proporcional a frequência (contagem) daquele nível.

```
gf_bar( ~ syll, data = magAds)
```



- O gráfico de barras também pode ser dividido por grupo:

```
gf_bar( ~ syll | GROUP, data = magAds)
```



3.2 Os dados de Ericksen

- Descrição dos dados: Ericksen 1980 U.S. Census Undercount.
- Este conjunto de dados contém as seguintes variáveis:
  - `minority`: Percentual de negros ou hispânicos.
  - `crime`: Taxa de crimes graves por 1000 indivíduos na população.
  - `poverty`: Percentual de pobres.
  - `language`: Percentual com dificuldade em falar ou escrever Inglês.
  - `highschool`: Percentual com idade igual ou superior a 25 anos que não terminou o ensino médio.
  - `housing`: Percentual de habitação em pequenos edifícios de unidades múltiplas.
  - `city`: Um fator com níveis: `city` (cidade principal) ou `state` (estado or estado-resto).
  - `conventional`: Percentual de domicílios contados por enumeração pessoal convencional.
  - `undercount`: Estimativa preliminar de subentendimento percentual.
- Os dados de Ericksen têm 66 linhas/observações e 9 colunas/variáveis.
- As observações são medidas em 16 grandes cidades, as partes restantes dos estados em que essas cidades estão localizadas, e os outros estados dos EUA.

```
Ericksen <- read.delim("C:/Users/fsabino/Desktop/Codes/papers/Introductory_Stat_I/notebook/datasets_Eri
head(Ericksen)
```

```
##   minority crime poverty language highschool housing  city conventional
## 1     26.1    49      19      0.2         44     7.6 state            0
## 2      5.7    62      11      1.7         18    23.6 state          100
## 3     18.9    81      13      3.2         28     8.1 state           18
## 4     16.9    38      19      0.2         44     7.0 state            0
## 5     24.3    73      10      5.0         26    11.8 state            4
```

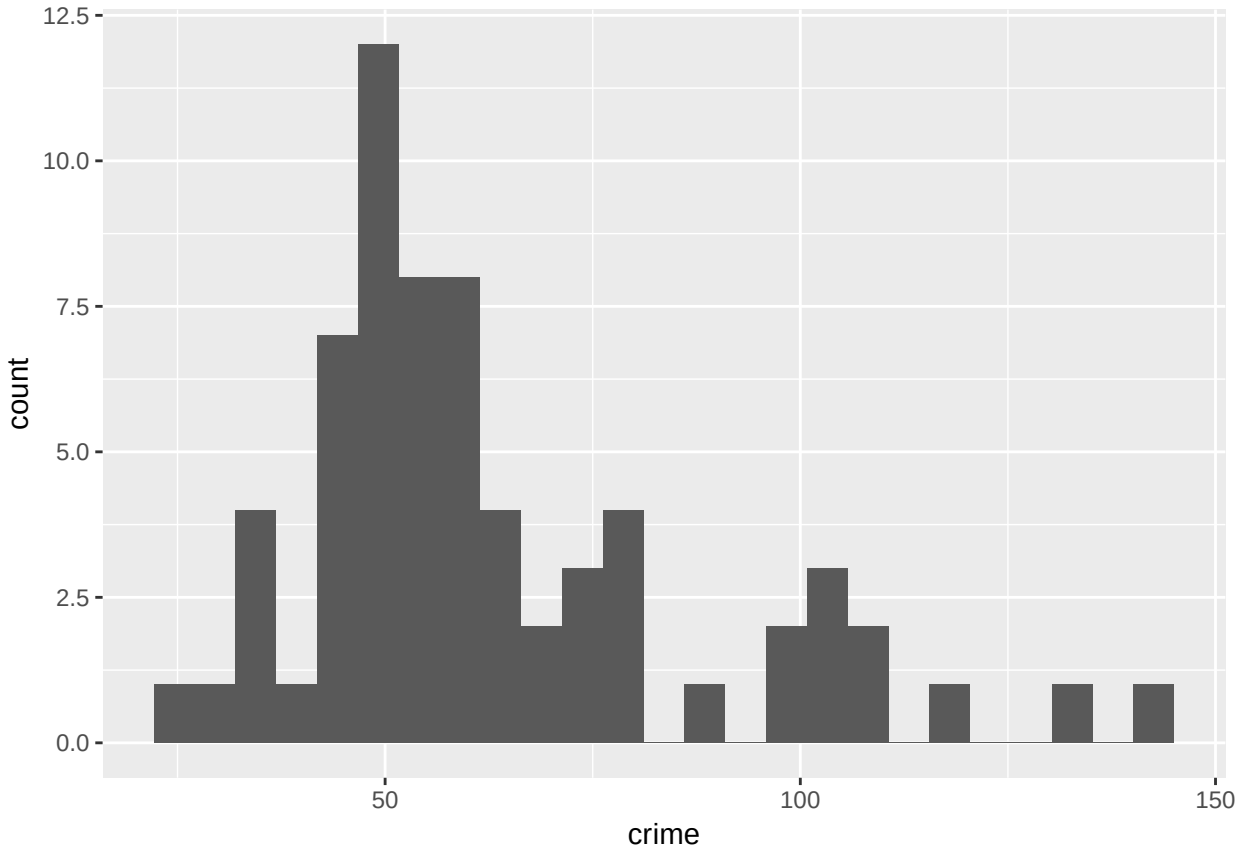
```
## 6     15.2    73      10      1.2         21     9.2 state           19
##   undercount
## 1      -0.04
## 2       3.35
## 3       2.48
## 4      -0.74
## 5       3.60
## 6       1.34
```

- Quer fazer um histograma para a taxa de criminalidade - como?

3.3 Histograma (usado para variáveis quantitativas)

- Como fazer um histograma para alguma variável `x`:
  - Divida o intervalo do valor mínimo de `x` para o valor máximo de`x` em um número apropriado de sub-intervalos de tamanho igual.
  - Desenhe uma caixa em cada sub-intervalo, sendo a altura proporcional ao número de observações no subintervalo.
- Histograma de taxas de criminalidade para os dados de Ericksen

```
gf_histogram( ~ crime, data = Ericksen)
```



Questão: Explique como o histograma é construído.

4 Resumo de Variáveis Quantitativas

4.1 Medidas de centro dos dados (tendência central/posição): Média, Mediana e Moda

- Retornemos ao exemplo de anúncios da revista (`WDS` = número de palavras no anúncio). Uma série de resumos numéricos para `WDS` pode ser encontrada usando a função `favstats`:

```
favstats( ~ WDS, data = magAds)
```

```
##  min Q1 median  Q3 max mean sd  n missing
##   31 69     96 202 230  123 66 54       0
```

- Os valores observados da variável `WDS` são $y_1 = 205$, $y_2 = 203, \ldots, y_n = 208$, onde existe um total de $n = 54$ valores. Conforme definido anteriormente, isso constitui uma **amostra**.
- 123 é a **média** da amostra, que é calculada por

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

  Nós chamamos $\bar{y}$ de **média amostral**.

```
mean(y)
```

```
## [1] 123
```

- A média é o ponto de equilíbro dos dados.
- A média é sensível a valores extremos.
- **mediana** = 96 é o percentil 50, i.e. o valor que divide a amostra em 2 grupos de igual tamanho.

```
median(y)
```

```
## [1] 96
```

```
* Veja como calcular a mediana (e qualquer outro percentil) na página 75 do livro "Estatística Aplicada
```

- A **mediana** é robusta a valores extremos.
- É uma medida mais apropriada quando trabalhamos com dados assimétricos.
- Uma propriedade importante da **média** e da **mediana** é que elas têm a mesma unidade de medida que as observações.
- **moda** = 208 é o valor mais frequente do banco de dados.

```
names(sort(-table(y)))[1]
```

```
## [1] "208"
```

- Exercício: Faça o exercício 1 da página 24 do livro Estatística (Costa Neto). A resposta está na página 258.

4.2 Medidas de variabilidade: amplitude, amplitude interquartílica, variância, desvio padrão, e coeficiente de variação

- Nós queremos saber "Quanto as observações estão desviadas do seu valor central?"
  - Ao olhar os dados e gráficos podemos ter uma sensação disto.

  - Porém, é comum estarmos interessados em um número para que possamos comparar as distribuições amostrais.
- **Amplitude** é a diferença entre o maior (máximo) e o menor (mínimo) valor.
  - Ela só usa dois valores para o seu cálculo, isto é, não leva todos em consideração.
  - Como trabalhamos com uma amostra, a amplitude que encontraremos será a amostral, isto é, em geral, temos uma subestimativa da verdadeira amplitude.
- A **amplitude interquartílica** é a diferença entre os valores do terceiro quartil e do primeiro quartil, isto é, $Q_3 - Q_1$.
  - Ela utiliza 50% dos valores para o seu cálculo.
- The **variância (empírica)** é a média dos desvios quadrados em relação à média:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

- **sd = desvio padrão** $= s = \sqrt{s^2}$.
- Nota: Por exemplo, se as observações são medidas em metros, a unidade de medida da **variância** será metro^2 o que usualmente dificulta a interpretação. Por outro lado, o **desvio padrão** tem a mesma unidade de medida das observações.
- O **coeficiente de variação** (CV) é uma medida adimensional que serve para comparar a variabilidade de variáveis medidas em diferentes unidade de medida ou cujas médias e desvios-padrão sejam muito diferentes (mesma unidade de medida e grandezas muito diferentes). Portanto, o **CV** é uma medida de variabilidade relativa, ao contrário das demais vistas que são medidas de variabilidade absolutas. Define-se o CV como a razão entre o desvio-padrão e a média (frequentemente é multiplicado por 100% para ser representado como uma variação percentual).

$$CV = \frac{s}{\bar{x}}$$

* Exercício: Faça o exercício 4 da página 32 do livro Estatística (Costa Neto). O objetivo do exercício é treinar o cálculo. A resposta está na página 258.

4.3 Cálculo da média, mediana, amplitude interquartílica e desvio-padrão usando a função favstats do pacote mosaic

- Medidas Resumo de `WDS`:

```
favstats( ~ WDS, data = magAds)
```

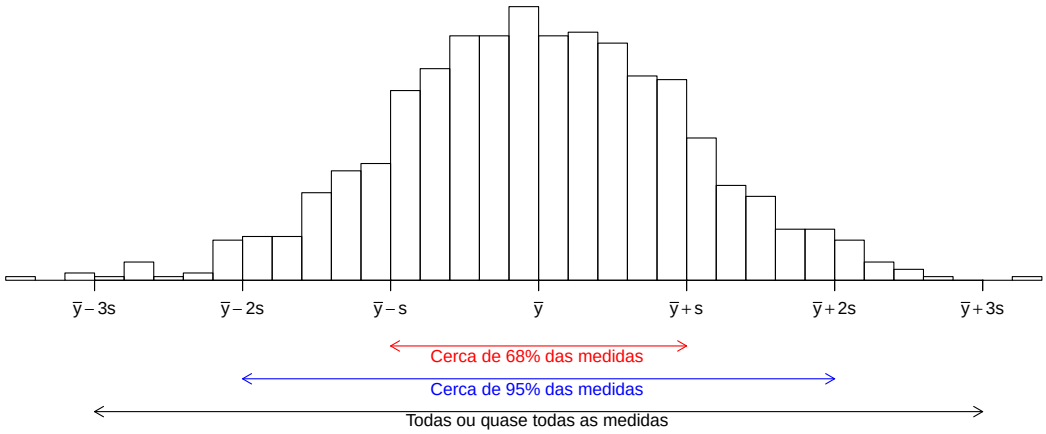
```
##  min Q1 median  Q3 max mean sd  n missing
##   31 69     96 202 230  123 66 54       0
```

Exercício: Interprete os resultados acima.

4.4 Uma palavra sobre terminologia

- **Desvio padrão**: uma medida de variabilidade de uma variável na amostra (ou população).
- **Erro padrão**: uma medida de variabilidade de uma estimativa (um particular valor de uma função da amostra). Por exemplo, uma medida de variabilidade da média amostral.

4.5 Uma regra empírica (veremos detalhes mais à frente)



Se o histograma com base na amostra parece uma função em forma de sino, então

- cerca de 68% das observações estão entre $\bar{y} - s$ e $\bar{y} + s$.
- acerca de 95% das observações estão entre $\bar{y} - 2s$ e $\bar{y} + 2s$.
- Todas ou quase todas as observações (99.7%) estão entre $\bar{y} - 3s$ e $\bar{y} + 3s$.

4.6 Percentis

- **O p-ésimo percentil** é um valor tal que pelo menos p% das observações são menores ou iguais a esse valor e pelo menos.
- Veja como calcular os percentis nas páginas 75-77 do livro texto.

4.7 Mediana, quartis e amplitude interquartílica

Recordando

```
favstats( ~ WDS, data = magAds)
```

```
##  min Q1 median  Q3 max mean sd  n missing
##   31 69     96 202 230  123 66 54       0
```

- 50-percentil = 96 é a **mediana** e é uma medida de tendência central/posição (centro dos dados).
- 0-percentil = 31 é o valor **mínimo**.
- 25-percentil = 69 é o **primeiro quartil** ou **quartil inferior** (Q1). Mediana dos 50% menores valores.
- 75-percentil = 201.5 é o **terceiro quartil** ou **quartil superior** (Q3). Mediana dos 50% maiores valores.
- 100-percentil = 230 é o valor **máximo**.
- **Amplitude Interquartílica (IQR)**: uma medida de variabilidade dada pela diferença entre o quartil superior e o quartil inferior: 201.5 - 69 = 132.5.

5 Mais gráficos

5.1 Box plots

Como desenhar um box plot:

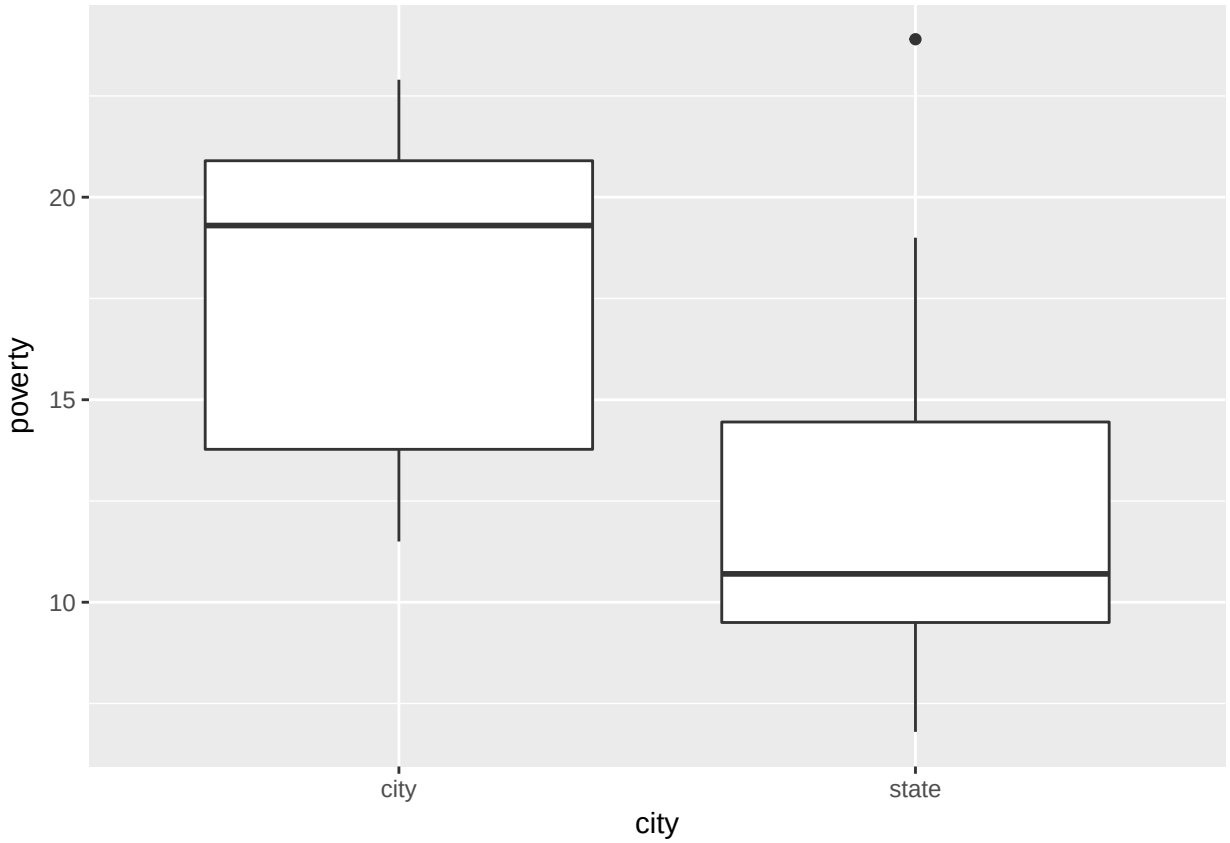
- Box:
  - Calcule a mediana, e os quartis inferior e superior.
  - Trace uma linha na mediana e desenhe uma caixa entre os quartis superior e inferior.
  - Calcule a amplitude interquartílica e a chame de IQR.
  - Calcule os seguintes valores:
    - L = quartil inferior - 1.5*IQR
    - U = quartil superior + 1.5*IQR
  - Desenhe uma linha ligando o quartil inferior até a menor medida que seja maior do que L.
  - Similarmente, desenhe uma linha ligando o quartil superior até a maior medida que seja inferior a U.
  - Regras de decisão
  - Max($X_[1]$, L)
  - Min($X_[n]$, U)
- Outliers: Observações com valor menor do que L ou maior do que U são desenhadas como círculos.

Nota: As caixas são fechadas (em inglês, as extremidades são chamadas de "Whiskers") no mínimo e no máximo das observações que não são consideradas outliers.

5.1.1 Boxplot para os dados de Ericksen

Boxplot das taxas de pobreza separadamente para cidados e estados (variável `city`):

```
gf_boxplot(poverty ~ city, data = Ericksen)
```

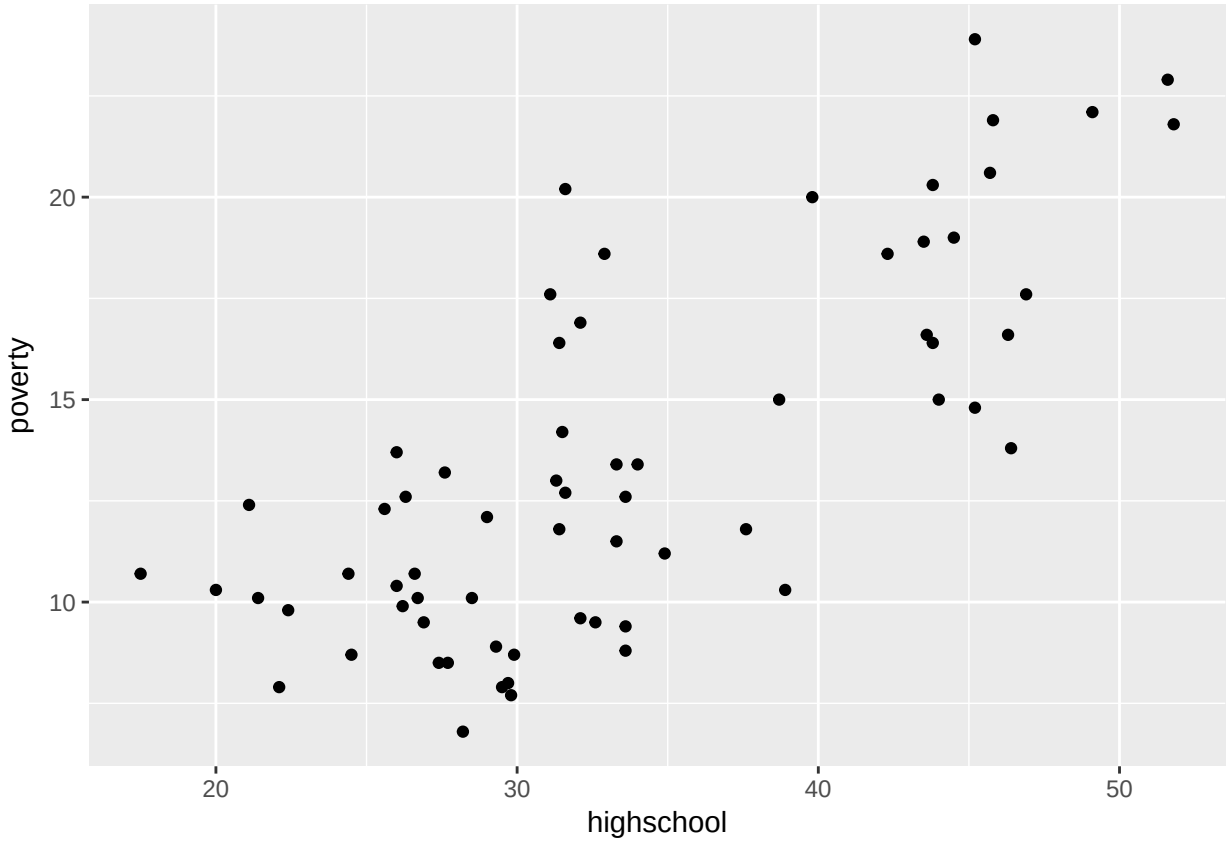


- Parece haver mais pobreza nas cidades.
- Um único estado difere notoriamente dos outros com alta taxa de pobreza.

5.2 2 variáveis quantitativas variables: Gráfico de dispersão ("Scatter plot")

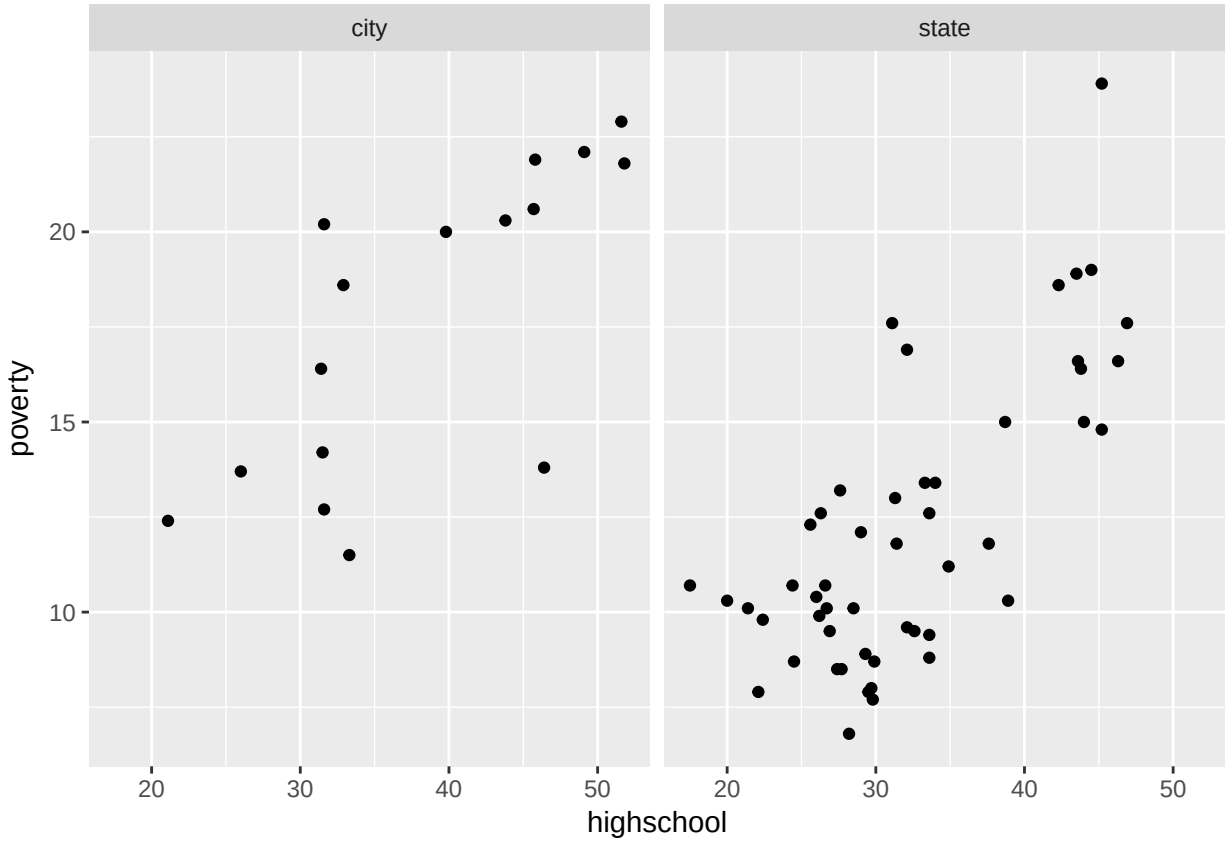
Para duas variáveis quantitativas, um gráfico frequentemente utilizado é o de dispersão:

```
gf_point(poverty ~ highschool, data = Ericksen)
```

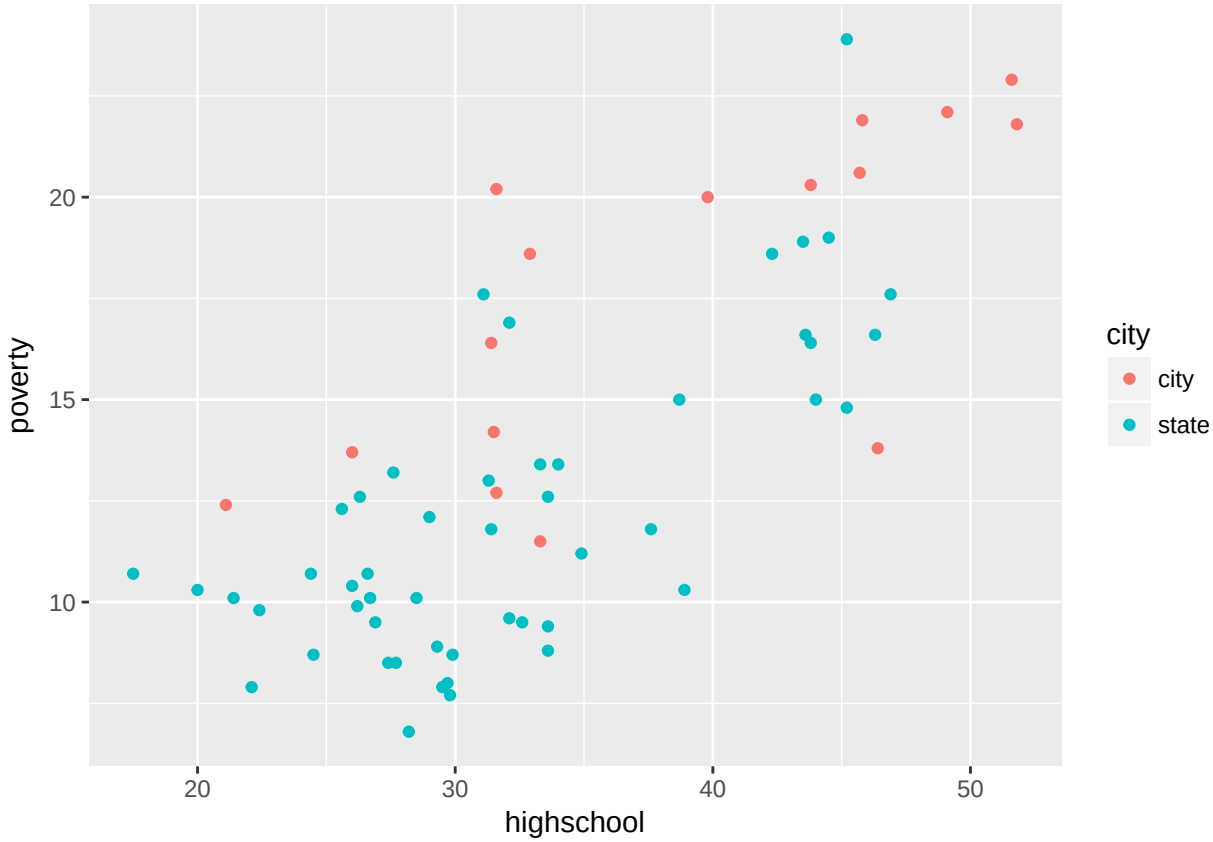


Isto pode ser colorido ou dividido de acordo com o valor de `city`:

```
gf_point(poverty ~ highschool | city, data = Ericksen)
```

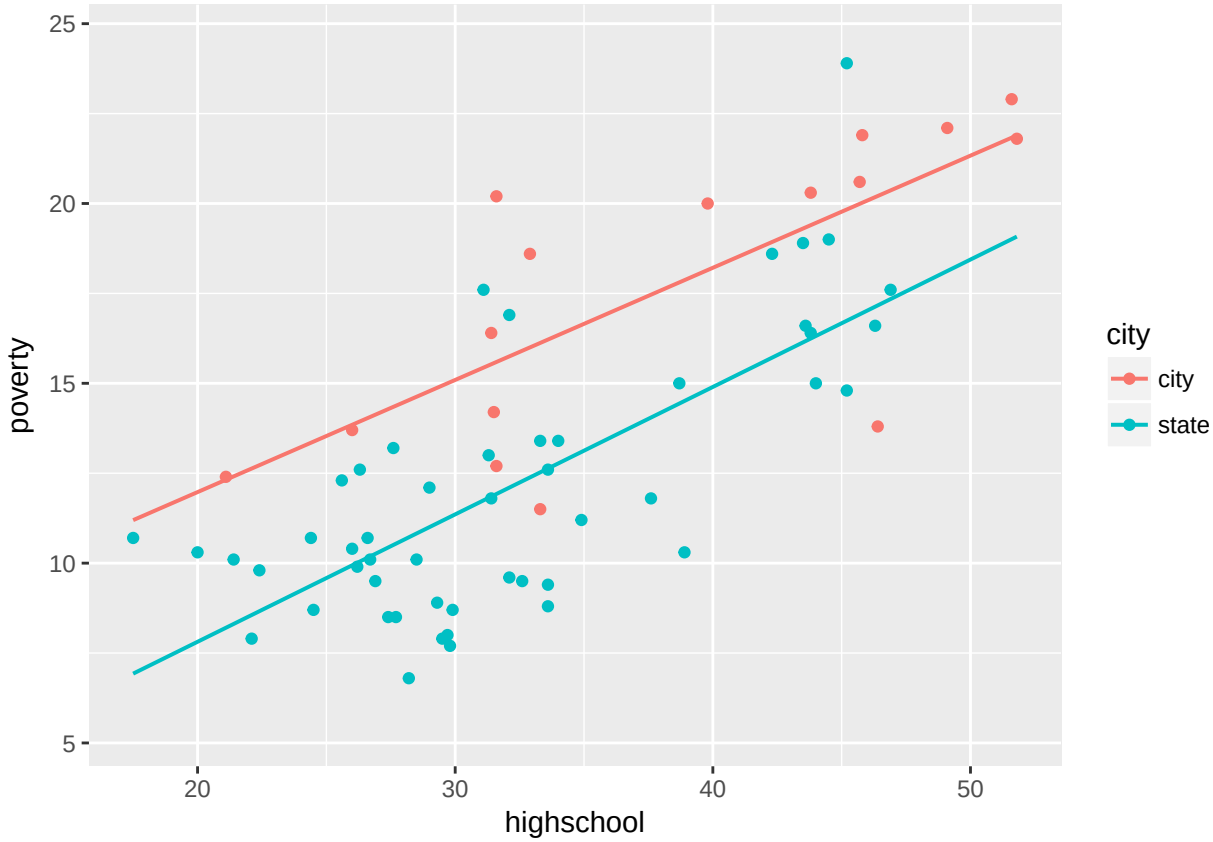



```
gf_point(poverty ~ highschool, col = ~city, data = Ericksen)
```



Se nos quisermos adicionar uma linha de regressão (uma equação da reta neste caso) nós podemos usar as funções abaixo:

```
gf_point(poverty ~ highschool, col = ~city, data = Ericksen) %>% gf_lm()
```



5.3 Assimetria e Curtose

- O conteúdo pode ser estudado nas páginas 30-31 do livro "Estatística" (Costa Neto).

5.4 Covariância e Correlação

```
* O conteúdo pode ser estudado nas páginas 98-104 do livro "Estatística Aplicada à Administração e Econ
```

6 Apêndice

6.1 Recodificando variáveis

- A função `factor` converterá diretamente um vetor em uma variável qualitativa (escala nominal). Por exemplo:

```
head(magAds$GROUP)
```

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

```
class(magAds$GROUP)
```

```
## [1] "integer"
```

```
f <- factor(magAds$GROUP)
class(f)
```

```
## [1] "factor"
```

```
# magAds$GROUP <- f
# head(magAds$GROUP)
```

- Desta forma, os números são substituídos por rótulos mais informativos descrevendo o nível educacional.

7 Apontar e clicar no gráfico

7.1 `mplot`

- Se os pacotes `mosaic` e `manipulate` forem instalados e estiverem carregados, nós podemos construir gráficos usando a função `mplot` simplesmente apontando e clicando.
- Usando `mplot` você pode fazer alterações pressionando o botão de configurações (uma roda dentada) no canto superior esquerdo da janela gráfica.

```
mplot(Ericksen)
```

- No final, você pode pressionar "Mostrar expressão" (Show expression) para obter o código.