Notes: MS 204 Chapter 1 part II

Overview

- Numerical data
- Univariate visualizations for numerical data

Numerical data

Univariate: For random variable X, we set out to collect X_1, X_2, ... X_n. The observed data is defined as $X_1 = x_1$, $X_2 = x_2$, ... $X_n = x_n$

Center

Shape

Spread

```
library(tidyverse)
library(gapminder)
summary(mtcars$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.43   19.20   20.09   22.80   33.90
```

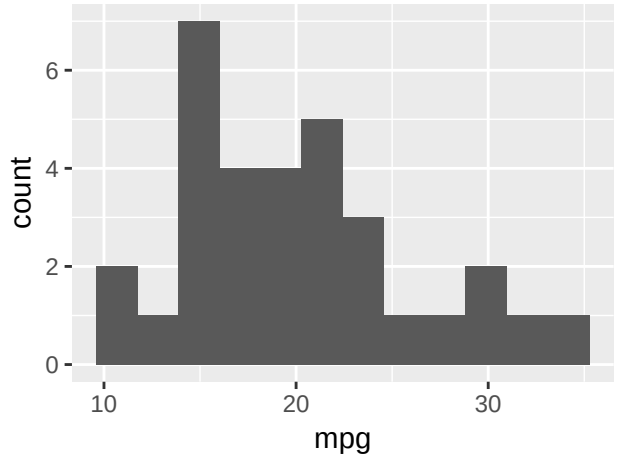
Population versus sample

Parameter versus statistic

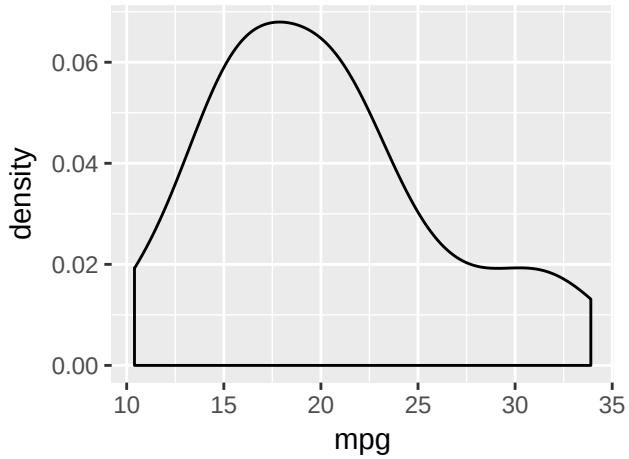
Aside: Sketch the expected distributions of (i) number of piercings, (ii) scores on an exam, (iii) IQ scores

Visualizing numerical data

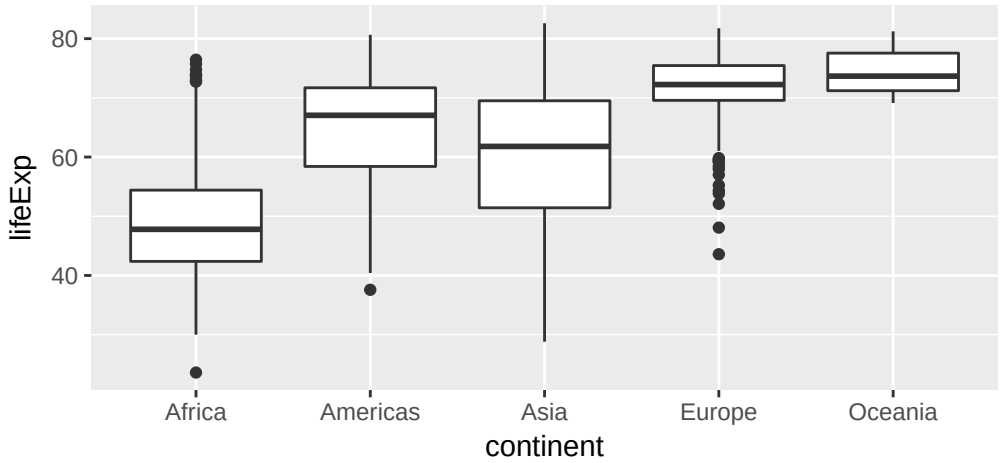
```
ggplot(mtcars, aes(x = mpg)) + geom_histogram(bins = 12)
```




```
ggplot(mtcars, aes(x = mpg)) + geom_density()
```




```
ggplot(gapminder, aes(x = continent, y = lifeExp)) + geom_boxplot()
```



Aside: Pluses and minuses of each of the above visualitions