

# Ch. 6: Comparing Two Means

# Navigation

## By Date

- March 31: start - end
- April 2: start - end
- April 7: start - end

## By Section

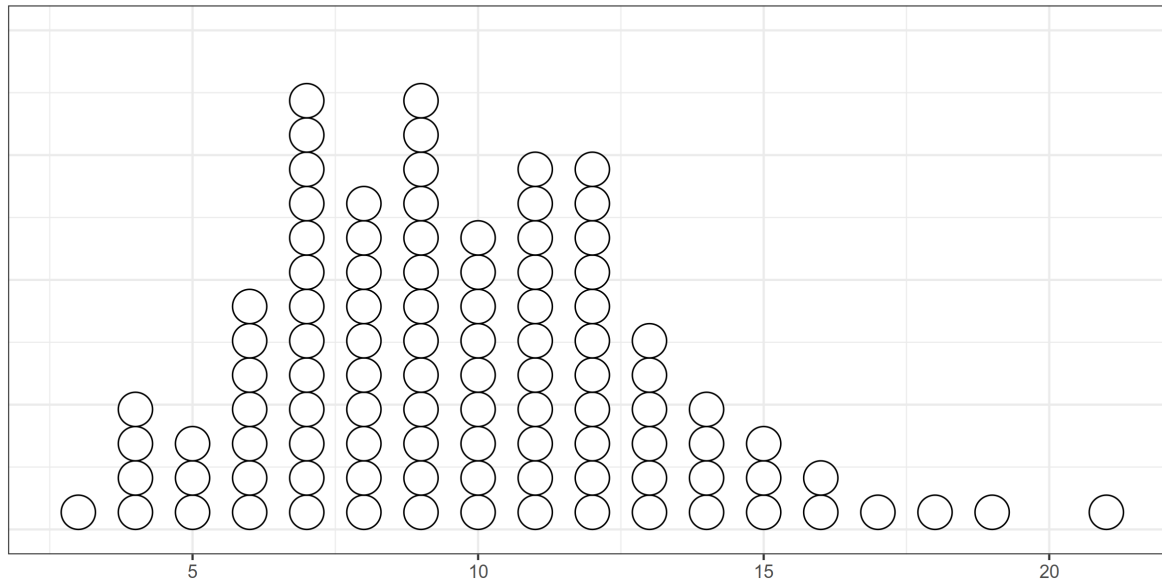
- 6.1: start - end
- 6.2: start - end
- 6.3: start - end

# 6.1: Comparing Two Groups

## Quantitative Response

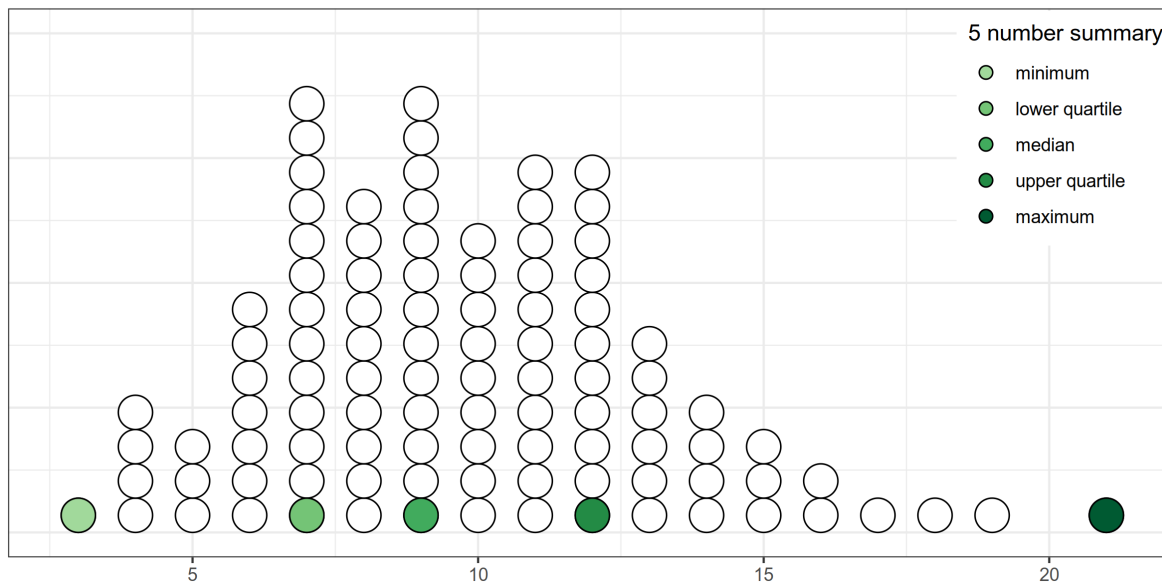
# Describing Distributions of Quantitative Data

- **5-number summary** - the minimum, lower quartile, median, upper quartile, and maximum of a set of data
  - **lower quartile** - 25% of the data lie below this value
  - **median** - 50% of the data lie below this value
  - **upper quartile** - 75% of the data lie below this value

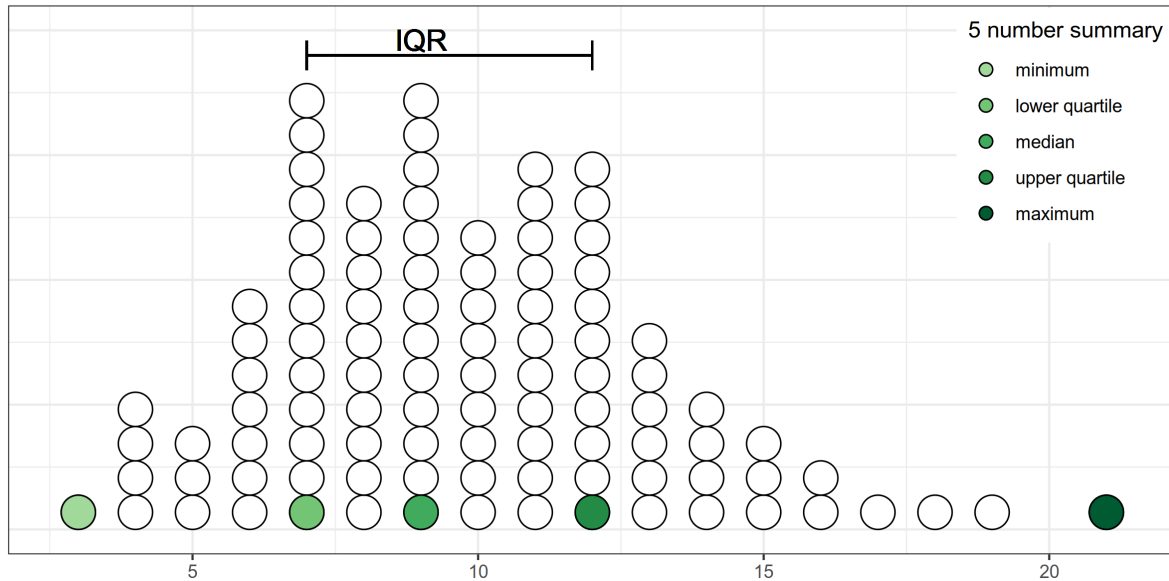


# Describing Distributions of Quantitative Data

- **5-number summary** - the minimum, lower quartile, median, upper quartile, and maximum of a set of data
  - **lower quartile** - 25% of the data lie below this value
  - **median** - 50% of the data lie below this value
  - **upper quartile** - 75% of the data lie below this value



# Describing Distributions of Quantitative Data



The distance between the two quartiles is called the **inter-quartile range**(IQR). The IQR is another measure of variability, along with the standard deviation.

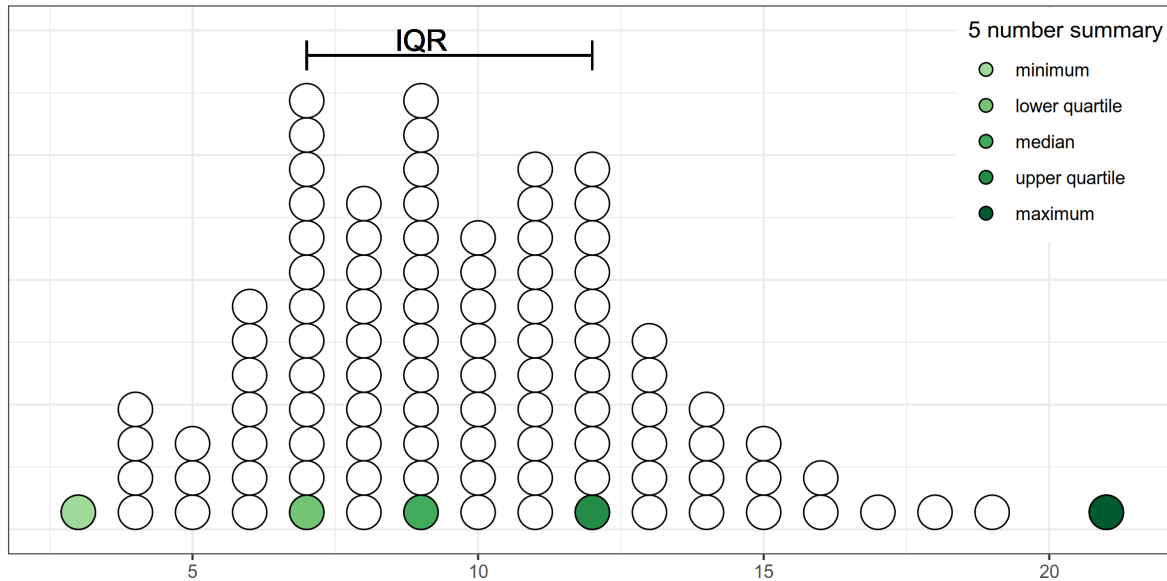
The IQR is resistant (or *robust*) to extreme values and skewness, unlike the standard deviation.

# Describing Distributions of Quantitative Data

A **boxplot** (or box-and-whisker plot) is a visual display of the 5-number summary.

- The box displays the middle 50% of the distribution and its width (the IQR) shows the spread of the bulk of the distribution.
- The 'whiskers' extend to the
  - smallest and largest values in the dataset
  - OR the values in the dataset that are within  $1.5 \times \text{IQR}$  away from the edges of the box.
    - Observations outside of the  $1.5 \times \text{IQR}$  range may be shown as dots and are outliers

# Describing Distributions of Quantitative Data



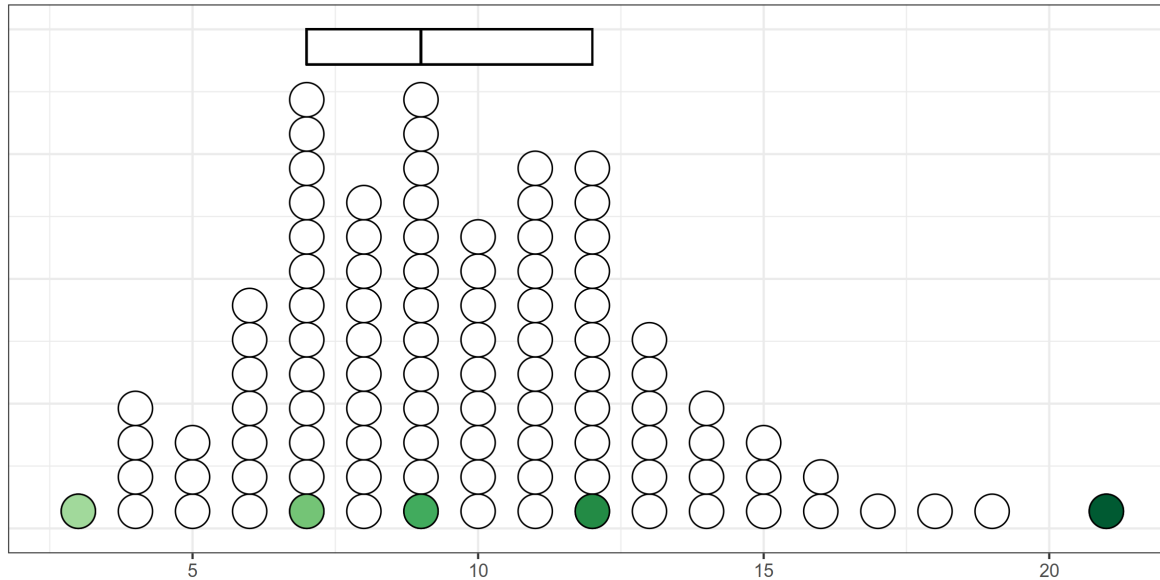
What are the following values?

- bottom whisker:
- lower quartile:
- median:
- upper quartile:
- top whisker:

Are there any outliers?



# Describing Distributions of Quantitative Data

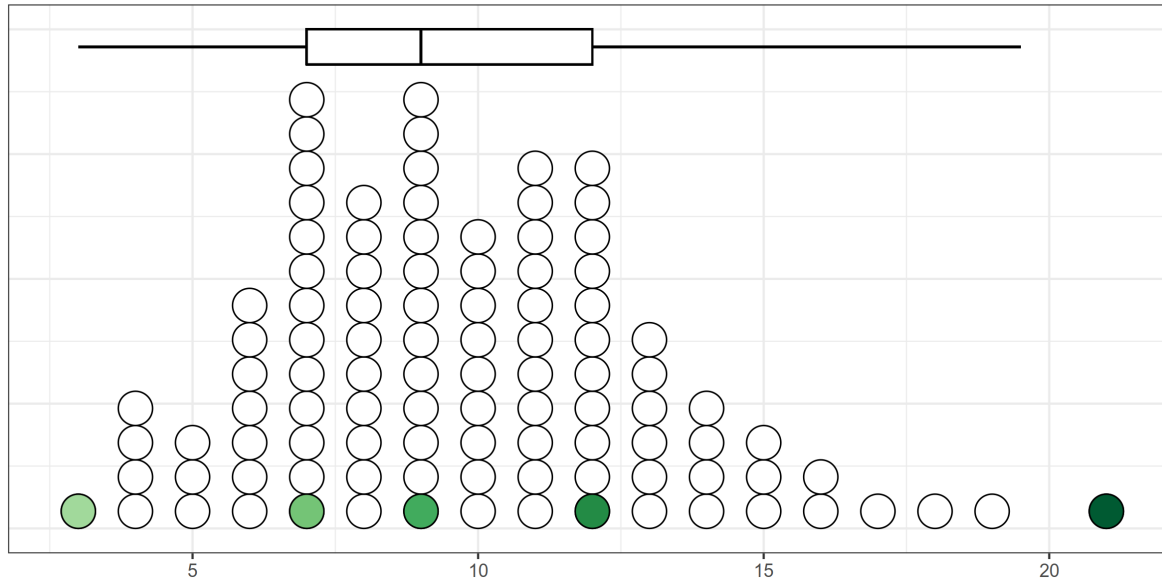


What are the following values?

- bottom whisker:
- lower quartile: 7
- median: 9
- upper quartile: 12
- top whisker:

Are there any outliers?

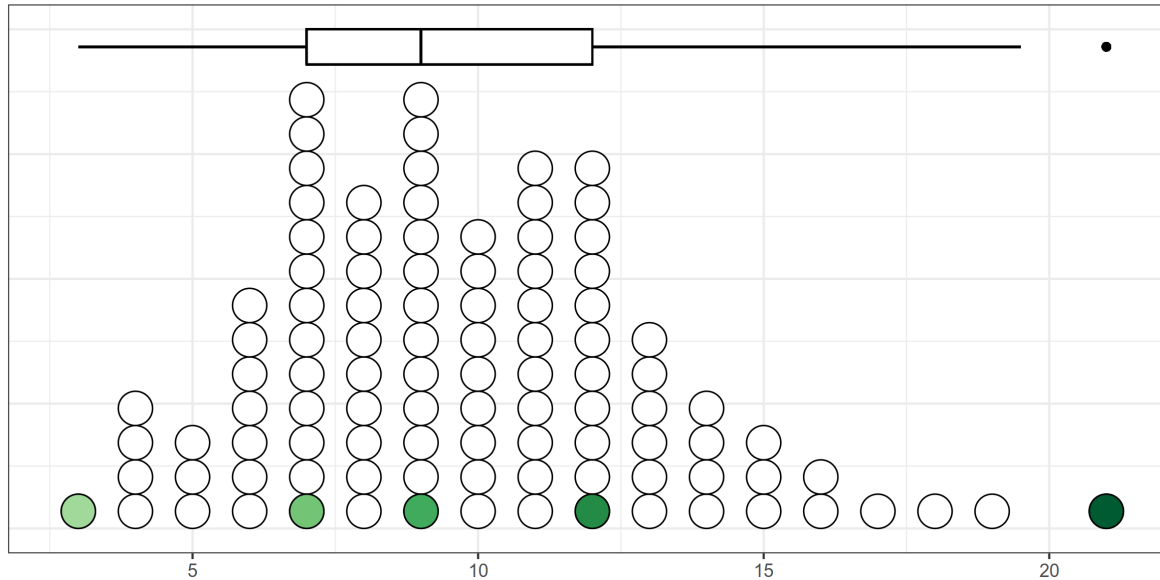
# Describing Distributions of Quantitative Data



- bottom whisker:  $1.5 \times \text{IQR}$  below lower quartile, or smallest observation  
 $7 - 1.5(12-7) = -0.5$  or 3  
bottom whisker: 3
- top whisker:  $1.5 \times \text{IQR}$  above upper quartile, or largest observation  
 $12 + 1.5(12-7) = 19.5$  or 21  
top whisker: 19.5

Are there any outliers?

# Describing Distributions of Quantitative Data



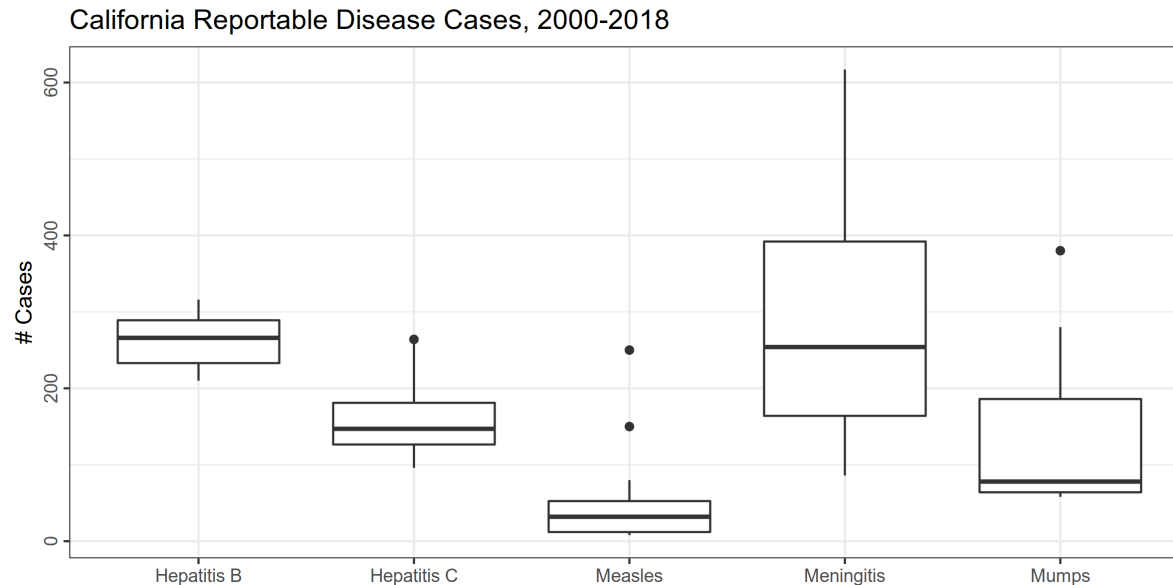
What are the following values?

- bottom whisker: 3
- lower quartile: 7
- median: 9
- upper quartile: 12
- top whisker: 19.5

Are there any outliers? Yes, at 21

# Describing Distributions of Quantitative Data

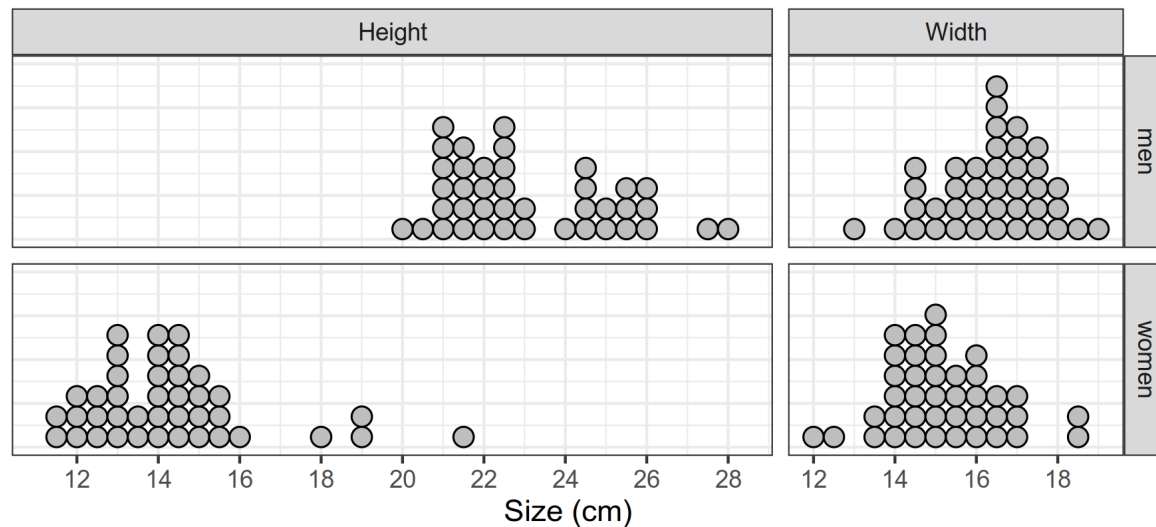
Boxplots make it easy to compare distributions of different groups.



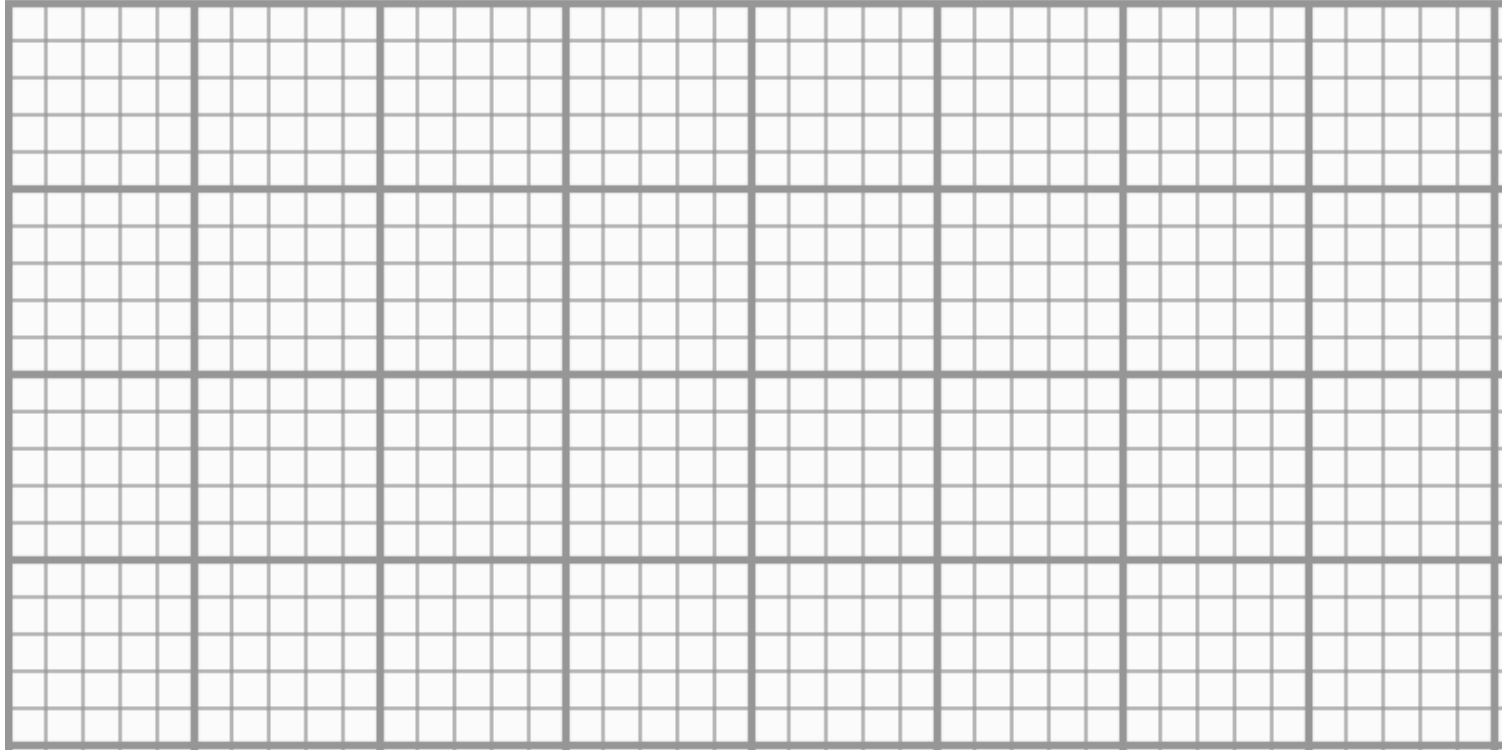
# In-Class Practice: Jean Pocket Sizes

Using the graph paper on the next page, create boxplots for each of the dimensions in the chart below.

Front Pocket Maximum



# In-Class Practice: Jean Pocket Sizes



What do you conclude about the distributions of front pocket maximum dimensions when comparing Men's jeans to Women's jeans?

# Exploration 6.1: Haircut Prices

Work through Exploration 6.1A in groups. Upload your answers; you must complete at least 1-13.

If you complete 14-19, you may get extra credit.

## 6.2: Comparing Two Means: Simulation Based Approach



# 3S strategy

1. Statistic
  2. Simulation
  3. Strength of Evidence
- 

Our hypotheses are going to concern  $\mu_1$  and  $\mu_2$ ... specifically, the value of  $\mu_1 - \mu_2$ .

We are interested in the long run (e.g. data collected forever) value of the difference between the group means

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 - \mu_2 \{ \neq, >, < \} 0$$

# Simulation

Group	Value	Simulated Group
1	a	1
1	b	2
1	c	1
2	d	1
2	e	2
2	f	2

Assign new group labels to existing values

Think about sorting labeled index cards into stacks for each group.

# Simulation

# Simulation

# Simulation

# Simulation

# Simulation

Null hypothesis: Mean \_\_\_\_\_ is the same in both groups

One repetition: Rerandomizing groups for each \_\_\_\_\_

Statistic: Difference in two group means

# Simulation

Null hypothesis: Mean measurement is the same in both groups

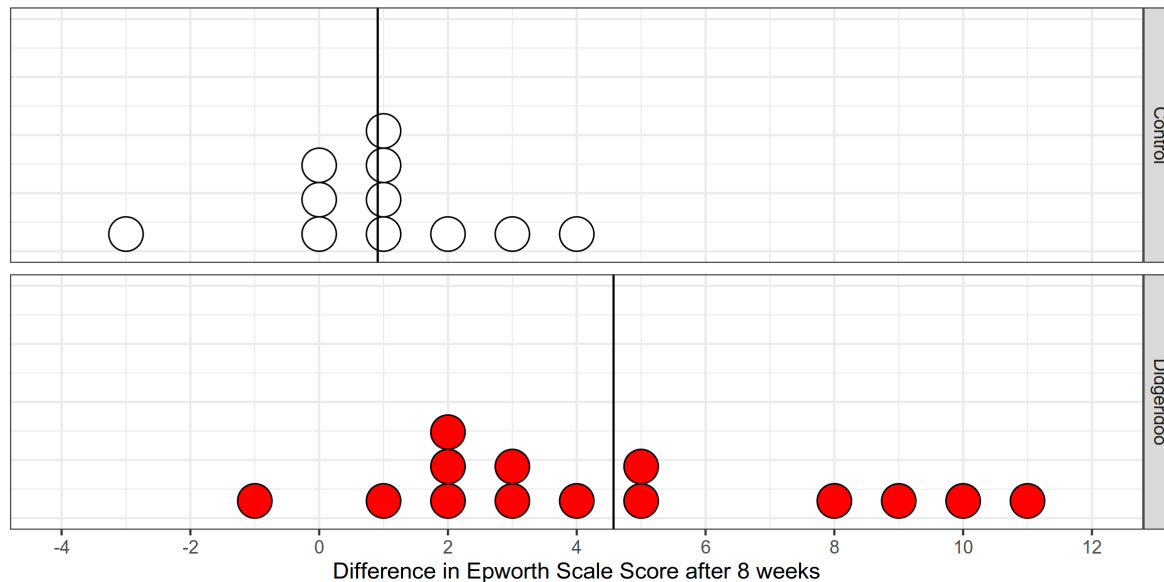
One repetition: Rerandomizing groups for each measurement

Statistic: Difference in two group means



# Example - Didgeridoo playing and sleep apnea

Scientists examined whether Didgeridoo lessons reduced the symptoms of sleep apnea as measured by the Epworth scale (0-24). Individuals were assessed for sleep apnea at the beginning and end of an 8-week period; the difference in scores was recorded. 14 individuals were randomly assigned to receive Didgeridoo lessons, while 11 did not receive any intervention.



# Example - Didgeridoo playing and sleep apnea

$$H_0 : \mu_D - \mu_C = 0$$

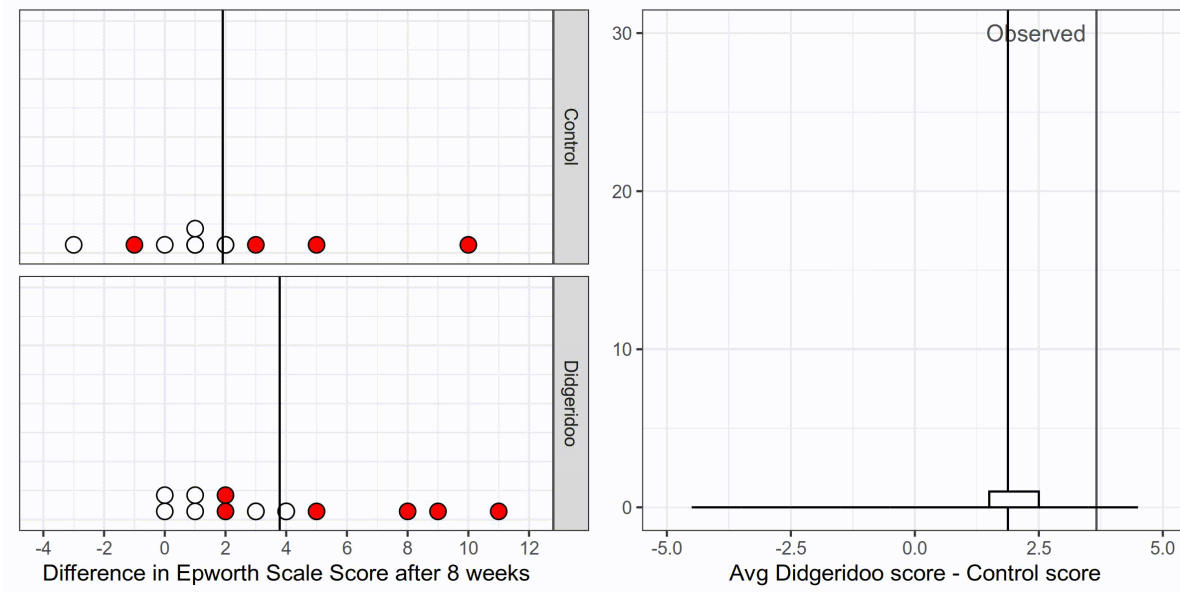
$$H_A : \mu_D - \mu_C > 0$$

In words:

- $H_0$ : The difference in the average change in Epworth Scale score between those who received Didgeridoo lessons and those that did not is 0
- $H_A$ : Those who received Didgeridoo lessons will have a higher average change in Epworth Scale score than those who did not.

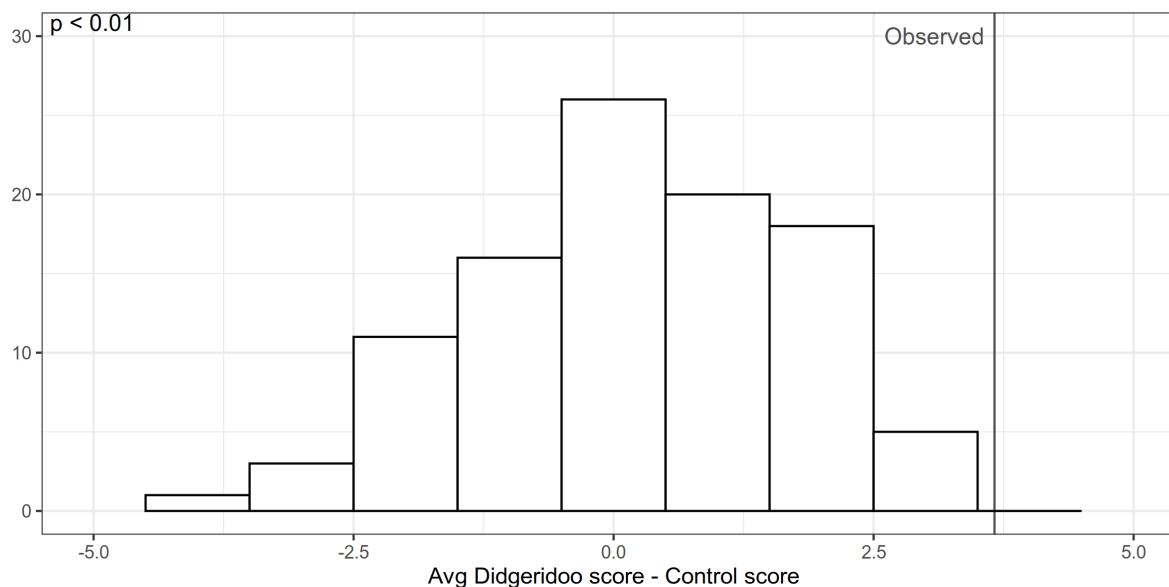
Because there is random assignment, if we conclude  $H_A$ , we can say the change is due to the Didgeridoo lessons.

# Example - Didgeridoo playing and sleep apnea



During each simulation, the group labels are shuffled and the difference in the means is calculated. The resulting simulation statistic distribution is used to calculate the p-value.

# Example - Didgeridoo playing and sleep apnea



## Interpretation:

With  $p < 0.01$ , we have very strong evidence against the null hypothesis that the two groups experienced equivalent reductions in Epworth Scale scores. Therefore, we reject  $H_0$  and conclude that it is more plausible that the group who received Didgeridoo lessons had a higher average reduction in Epworth Scale score than the control group. There is evidence that Didgeridoo lessons reduce the symptoms of sleep apnea.

# Exploration 6.2

Complete Exploration 6.2, questions 1 - 15 in your groups.

## 6.3: Comparing Two Means: Theory-Based Approach

## Example 6.3: Breastfeeding and Intelligence

An article published in the journal *Pediatrics* (1999) studied whether and how children who were breastfed during infancy differed from those who weren't breastfed. The study involved 323 white children recruited at birth in 1980-1981 from four Western Michigan Hospitals. After some initial exploration, the researchers deemed the participants in the study were representative of the community in terms of social class, maternal education, age, marital status, and sex of infant.

- Is this an experiment or an observational study?
- Is this a random sample?
- Is this a representative sample?
- Can we make causal conclusions (e.g. breastfeeding causes increased intelligence)

## Example 6.3: Breastfeeding and Intelligence

An article published in the journal Pediatrics (1999) studied whether and how children who were breastfed during infancy differed from those who weren't breastfed. The study involved 323 white children recruited at birth in 1980-1981 from four Western Michigan Hospitals. After some initial exploration, the researchers deemed the participants in the study were representative of the community in terms of social class, maternal education, age, marital status, and sex of infant.

- Is this ~~an experiment~~ or an observational study? **Investigators did not assign treatments (breastfeeding/not) to infants**
- Is this a random sample? **No - infants were not randomly selected - they were 'recruited'**
- Is this a representative sample? **Possibly, based on initial exploration by the researchers. If the population of interest is white infants, then maybe.**
- Can we make causal conclusions (e.g. breastfeeding causes increased intelligence) **No, we can't make causal conclusions, because there was no random assignment of treatments to observational units**



## Example 6.3: Breastfeeding and Intelligence

The researchers followed up with the children at age 4, assessing their abilities using the McCarthy Scales of Children's Abilities. The GCI (General Cognitive Index) was recorded for each child, along with whether or not the child was breastfed.

What should the researchers use as their hypotheses, in words?

## Example 6.3: Breastfeeding and Intelligence

The researchers followed up with the children at age 4, assessing their abilities using the McCarthy Scales of Children's Abilities. The GCI (General Cognitive Index) was recorded for each child, along with whether or not the child was breastfed.

What should the researchers use as their hypotheses, in words?

**Null:** There is no underlying association between breastfeeding during infancy and GCI at age 4.

**Alt:** There is an association between breastfeeding during infancy and GCI at age 4.

## Example 6.3: Breastfeeding and Intelligence

The researchers followed up with the children at age 4, assessing their abilities using the McCarthy Scales of Children's Abilities. The GCI (General Cognitive Index) was recorded for each child, along with whether or not the child was breastfed.

What are the hypotheses in symbols?

## Example 6.3: Breastfeeding and Intelligence

The researchers followed up with the children at age 4, assessing their abilities using the McCarthy Scales of Children's Abilities. The GCI (General Cognitive Index) was recorded for each child, along with whether or not the child was breastfed.

What are the hypotheses in symbols?

$$H_0 : \mu_{BF} - \mu_{Not} = 0$$

$$H_A : \mu_{BF} - \mu_{Not} \neq 0$$

OR

$$H_0 : \mu_{BF} = \mu_{Not}$$

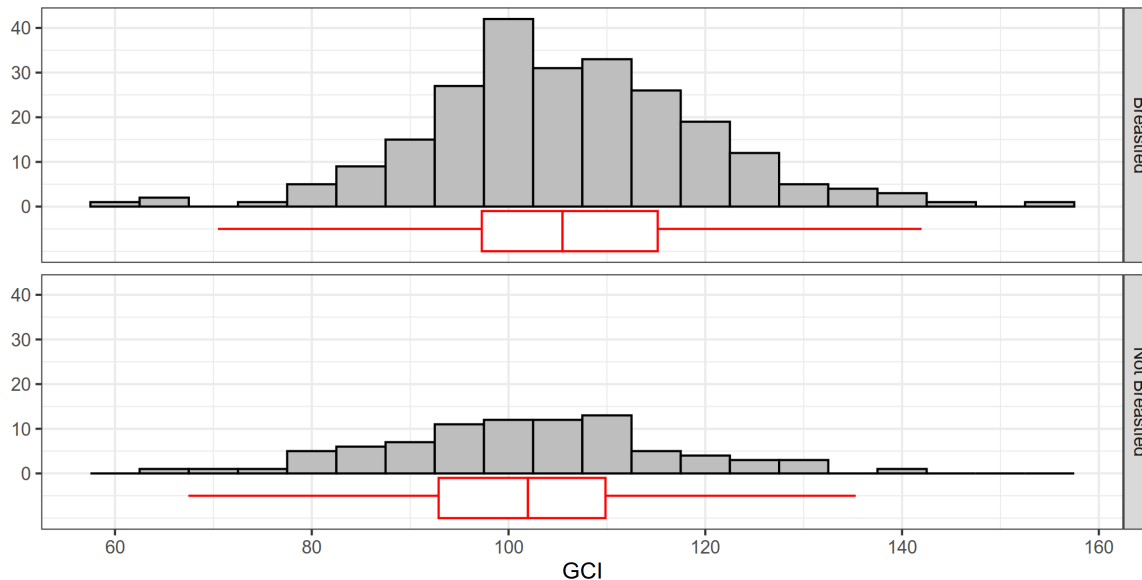
$$H_A : \mu_{BF} \neq \mu_{Not}$$

# Example 6.3: Breastfeeding and Intelligence

Summary Stats:

Group	n	Sample mean	Sample SD
Breastfed	237	105.30	14.50
Not Breastfed	85	100.90	14.00

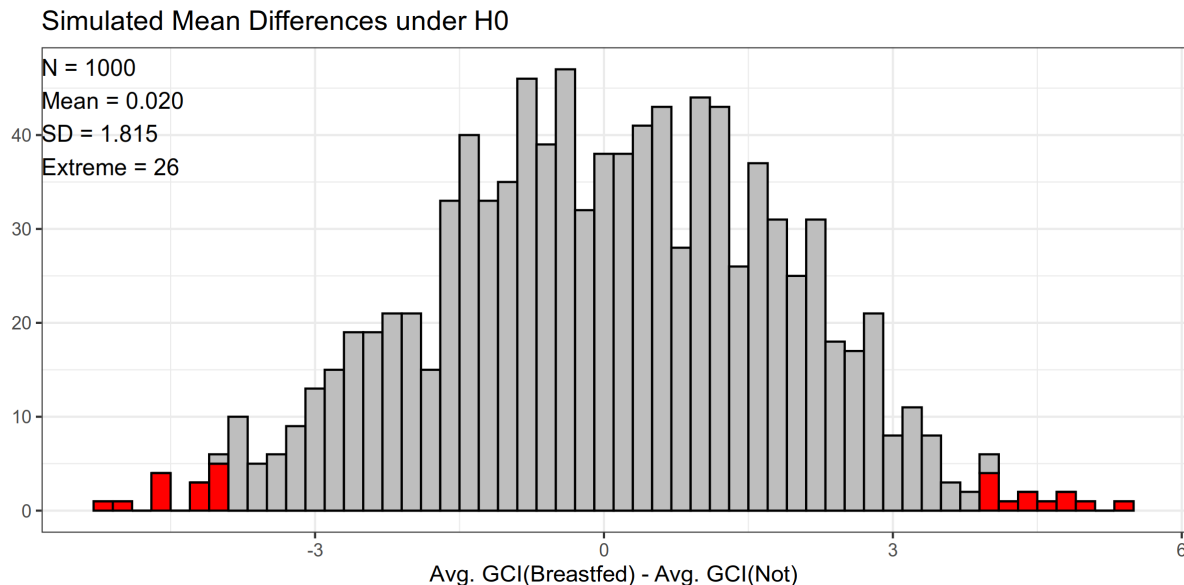
Simulated Data:



# Example 6.3: Breastfeeding and Intelligence

## CI Option 1: Theory-ish based Confidence Interval

- $\text{Statistic} \pm 2 \text{SD}(\text{Statistic})$   
SD of difference in means from simulation



$$3.961 \pm 2 (1.815) = (0.331, 3.650)$$

# Example 6.3: Breastfeeding and Intelligence

## CI Option 2: Theory based Confidence Interval

- Statistic  $\pm 2$  SD(Statistic)  
SD of difference in means from formula

$$\begin{aligned}SD(\bar{x}_1 - \bar{x}_2) &= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\&= \sqrt{\frac{14.5^2}{237} + \frac{14.0^2}{85}} \\&= \sqrt{0.887 + 2.306} \\&= 1.787\end{aligned}$$

$$3.961 \pm 2 (1.787) = (0.387, 3.594)$$

On a test, use Option 2 unless it explicitly says you may use the simulation standard error

# Example 6.3: Breastfeeding and Intelligence

A standardized statistic is thus

$$t = \frac{\text{statistic} - \text{hypothesized value}}{SE(\text{statistic})} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

---

## Validity Conditions

Either

- a symmetric distribution in both groups

OR

- at least 20 observations in each group and sample distributions that are not highly skewed

If the validity conditions hold, you can use the guideline that if  $t > 2$ , then the p-value is probably  $< 0.05$ .



# Factors Affecting the Width of a CI

- size of  $n_1$  and  $n_2$ 
  - If  $N = n_1 + n_2$  is constant, the difference between  $n_1$  and  $n_2$
  - the more balanced, the smaller the margin of error (all else equal)
  - Higher  $N$  = smaller margin of error
- variability of group 1 and/or group 2
- multiplier