**Lesson 2**

**Technically the height can be considered ordinal, which of the following is true:**

**A** It is more effective to consider heights to be numerical given the number of unique values we observe and even more we can potentially observe.

**What proportion of the data is between 69 and 72 inches (taller than 69 but shorter or equal to 72)? Hint: a logical operator and mean.**

library(dslabs)

data(heights)

x <- heights$height[heights$sex=="Male"]

mean(x>69 & x<=71)

**Lesson 3**

**Divide in group, can you do for female**

female <-heights$height[heights$sex=="Female"]

female\_percentiles <- quantile(female, seq(0.1, 0.9, 0.2))

**In which percentiles there is the maximum differences between the gender?**

**30, 50 and 70, it means that in the outlier we found less diversity**

**Divide in group:**

**can you compute the median and see if it corresponds to the boxplots? Where is the median for the females?** 64.98031

**And for the males?** 69

median(heights$height)

median(heights$height[heights$sex=="Male"])

median(heights$height[heights$sex=="Female"])

**Divide in group: Suppose all you know about the data is the average and the standard deviation. Use the normal approximation to estimate the proportion you just calculated.**

data(heights)

x <- heights$height[heights$sex=="Male"]

avg <- mean(x)

stdev <- sd(x)

pnorm(71, avg, stdev) - pnorm(69, avg, stdev)

**In group, Adjust the bin to have a ncie histogram**

ggplot(heights, aes(x=height)) + geom\_histogram(binwidth=1.5)

**Perform two different curves for sex, hint: use fill in the aes parameter**

ggplot(heights, aes(x=height, fill=sex)) + geom\_density(alpha=.3)