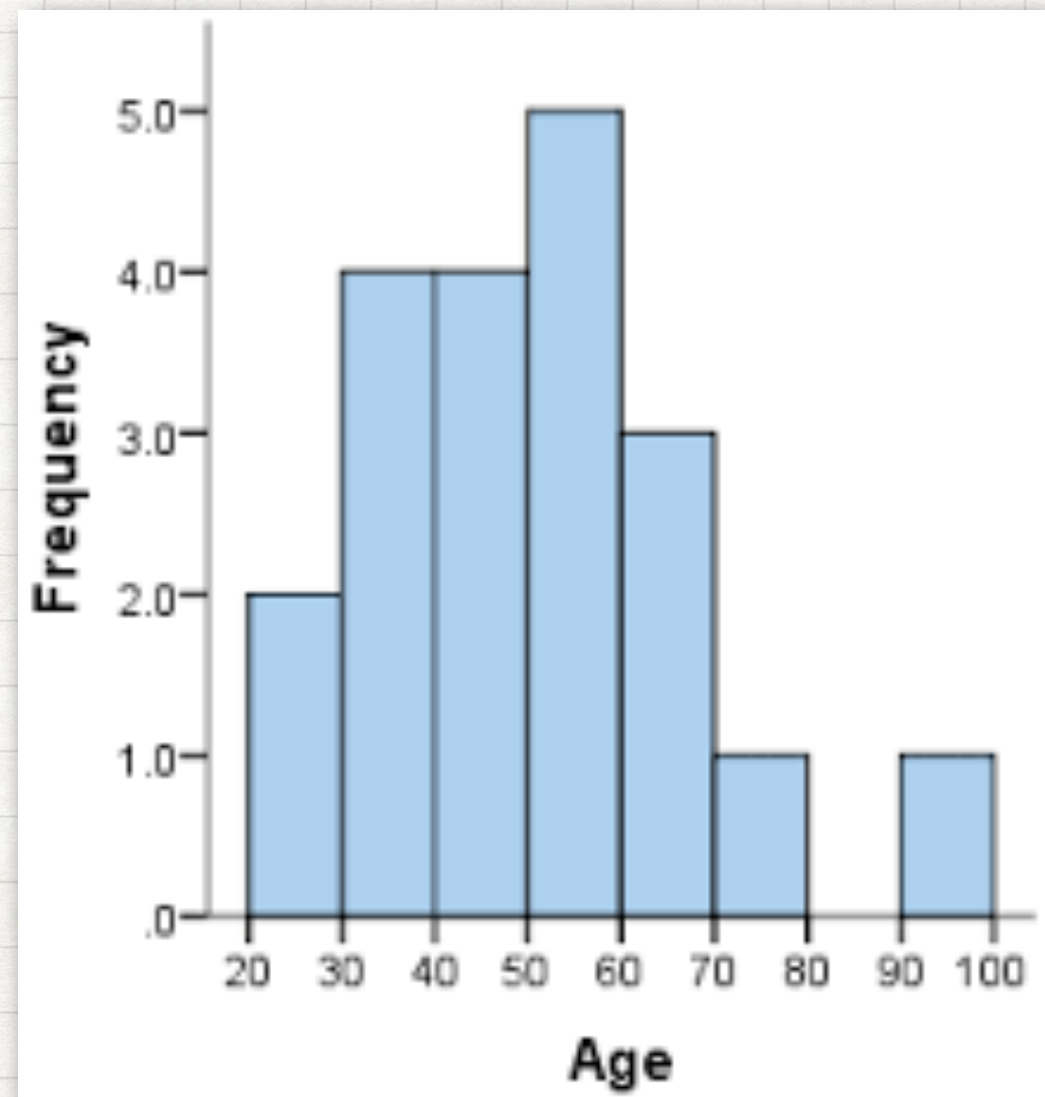


HISTOGRAM VIOLIN PLOTS BOX PLOTS

CHUMI, MARCO, SABRINA

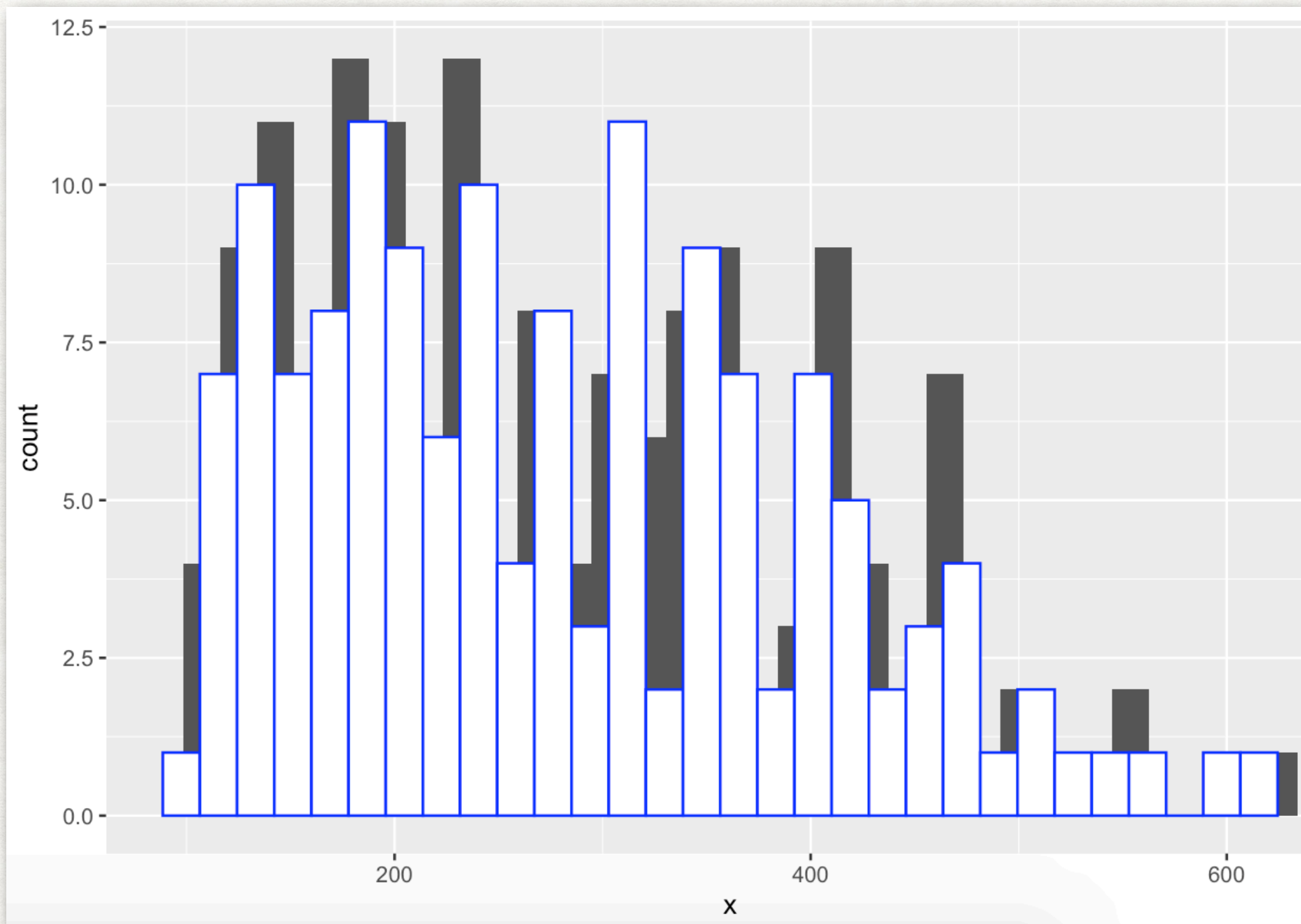
WHAT IS A HISTOGRAM

- visual representation of the distribution of a dataset
- it allows you to see large amount of the data is situated and where there is very little data to be found (Karjin Willems)
- x-axis, a y-axis many bars of different heights
- y axis= how frequently the values on the x-axis occur in the data
- x axis= bars group ranges of values or continuous categories



HOW TO CODE:

- `#How to plot a histogram`
- `AP <- as.data.frame(AirPassengers) #makes the AirPassengers data set into a dataframe`
- `H <- ggplot(AP, aes(x=x)) + geom_histogram() #creates a histogram with the Air passengers dataset`
- `H + geom_histogram(fill = "white", color = "black") #fills the graph in white with a black outline`
- `H + geom_histogram(binwidth = 20, fill = "white", color = "black", boundary = 35) #Changes the boundary and binwidth`
- `H + geom_histogram(color = "blue", bin = 20, fill = "white", color = "black", boundary = 35) #Changes the color of the histogram to blue`



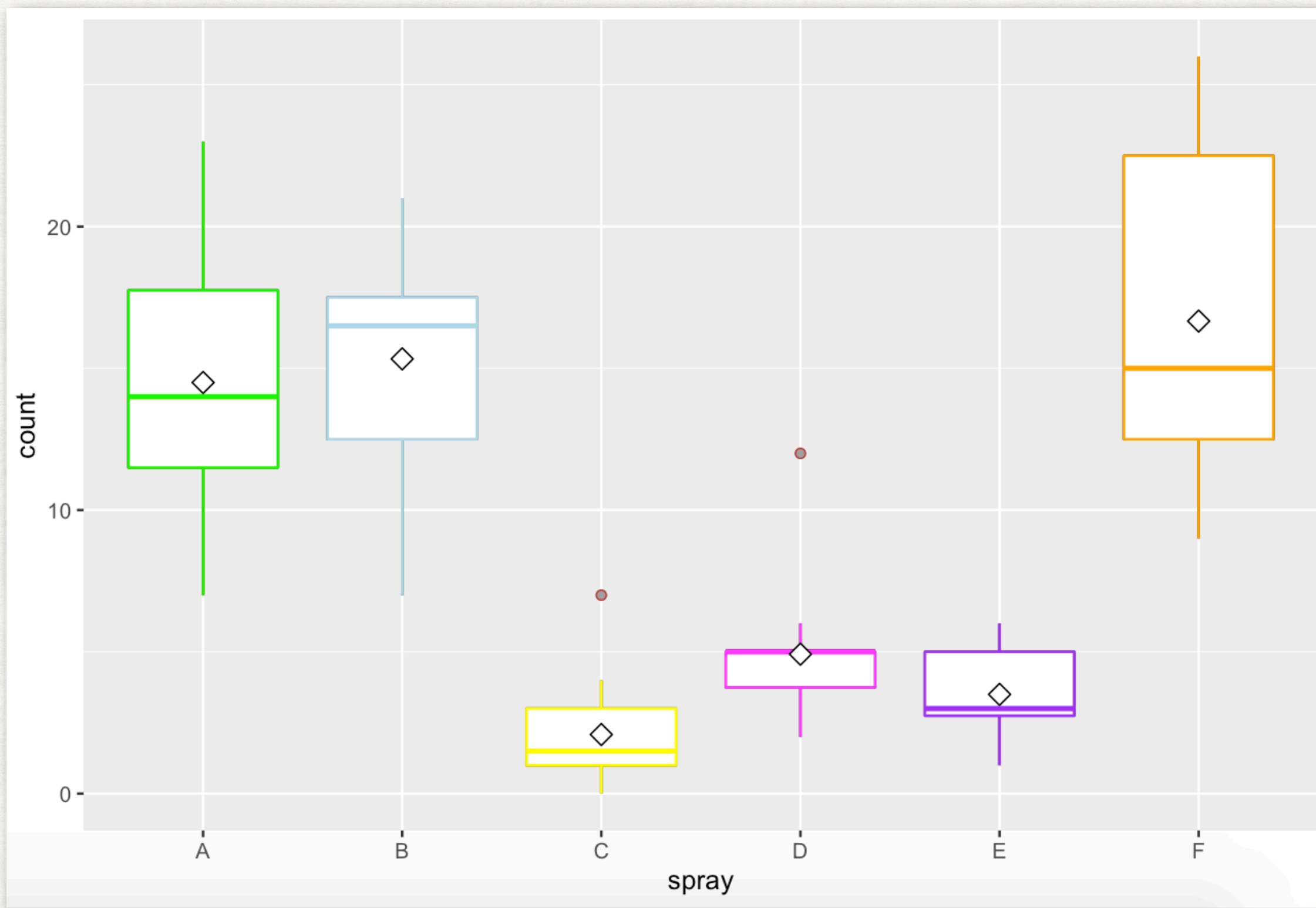
VIOLIN AND BOX PLOTS

- Box plots: how the values in data are spread out
- Minimum, Q1, Median, Q3, Maximum
- Advantage: takes up less space
- Violin plots: similar to box plots, also shows probability density of the data
- Includes Median of the data and a box showing the interquartile range

.

BOX PLOTS

- `B <- ggplot(InsectSprays, aes(x=spray, y=count)) + geom_boxplot()` #Creates a box plot from the dataset InsectSprays, spray vs. count, assigned to the variable B.
- `B + geom_boxplot(width= 0.5)` #Changes the width of the plot by a 0.5 scale.
- `B + geom_boxplot(outlier.size = 1.5)` #Changes the size of the point of the outlier of the box plot.
- `B + geom_boxplot(outlier.size = 1.5, outlier.shape = 21)` #Changes the point and shape of the outlier of the box plot.
- `B + geom_boxplot(outlier.size = 1.5, outlier.shape = 21, outlier.color = "Red", outlier.alpha = 1/2)` #Changes the point, shape, color, and alpha of the outlier of the box plot.
- `B + stat_summary(fun.y = "mean", geom = "point", shape= 23, size= 3, fill= "white")` #Adds a diamond, filled in white, of the median of each box plot
- `B + geom_boxplot(color = c("green", "light blue", "yellow", "magenta", "purple", "orange"), outlier.size = 1.5, outlier.shape = 21, outlier.color = "Red", outlier.alpha = 1/2) + stat_summary(fun.y = "mean", geom = "point", shape= 23, size= 3, fill= "white")` #changes the color of each box to a specified color, changes the size, point, color, and alpha of the outlier, and creates a median on each plot and assigns the value to a white diamond



VIOLIN PLOTS:

- `V <- ggplot(InsectSprays, aes(x=spray, y=count)) + geom_violin()` #Creates a violin plot from the dataset InsectSprays, spray vs. count, assigned to the variable V.
- `V + geom_violin(width= 0.1)` #Changes the width of the plot by a scale of 0.1.
- `V + geom_violin(width= 0.1, fill = "Black")` #changes the width of the plot by a scale of 0.1 and fills the central values in black
- `V + geom_violin(width= 0.1, fill = "Black", trim = FALSE)` #changes the width of the plot a scale of 0.1, fills the central values in black, and allows for the outliers to be shown
- `V + geom_violin(width= 0.1, fill = "Black", trim = FALSE) + stat_summary(fun.y = median, geom = "point", fill= "white", shape = 21, size = 2.5)` #adds a median value in gthe form of a white circle
- `V + geom_violin(scale= "count", width= 0.1, fill = "Black", trim = FALSE) + stat_summary(fun.y = median, geom = "point", fill= "white", shape = 21, size = 2.5)` #Scales the graph appropriately with the data
- `V + geom_violin(adjust= 0.5, scale= "count", width= 0.1, fill = "Black", trim = FALSE) + stat_summary(fun.y = median, geom = "point", fill= "white", shape = 21, size = 2.5)` #adjusts the smoothness of the curve
- `V + geom_violin(color = "pink", adjust= 0.5, scale= "count", width= 0.1) + stat_summary(fun.y = median, geom = "point", fill= "white", shape = 21, size = 2.5)` #changes the color of the middle line to pink

