Continuous y and Categorical x Data

Jessie, Isaac, Alexander, Manou

Continuous and Categorical geom_

-Include geom_bar, geom_boxplot, and geom_violin

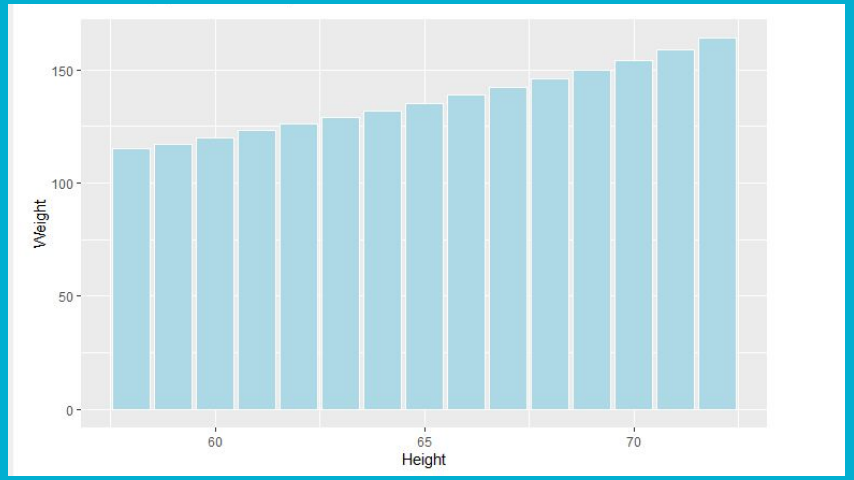
-Linked together by R data styles of numeric(continuous) and characters(categorical)

-Numeric(1.3, 2.5, 2.4, 3)

-Characters(“One”, “Two”, “Tree”, “Orange”)

geom_bar

- We used data(women), a compilation of average weight/height of American women



I used the color and fill to change the aesthetic code of my graph

Women$height and Women$weight were used to select the two sets of data we wanted

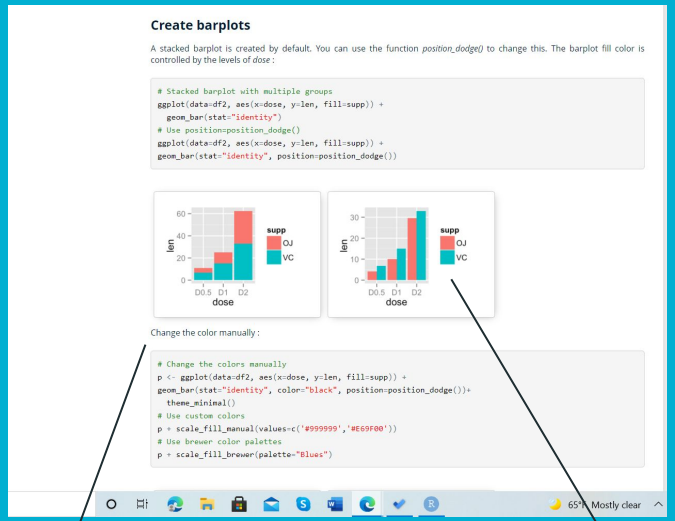
I used labs to name the x and y axis of the bar graph

```r
df <- data.frame(women$height, women$weight)
df
library(ggplot2)
p <- ggplot(data=df,aes(x=women$height, y=women$weight, fill=supp)) +
  geom_bar(stat="identity", color="white", fill="lightblue") +
  labs(y="Weight", x="Height", main="Average Weight and Height of American Women")
p
```

Different example of plotting a Geom_bar graph

- This would be another example of a bar graph and the code.

- This code is for creating a bar graph that contains multiple groups/data sets → a stacked bar graph

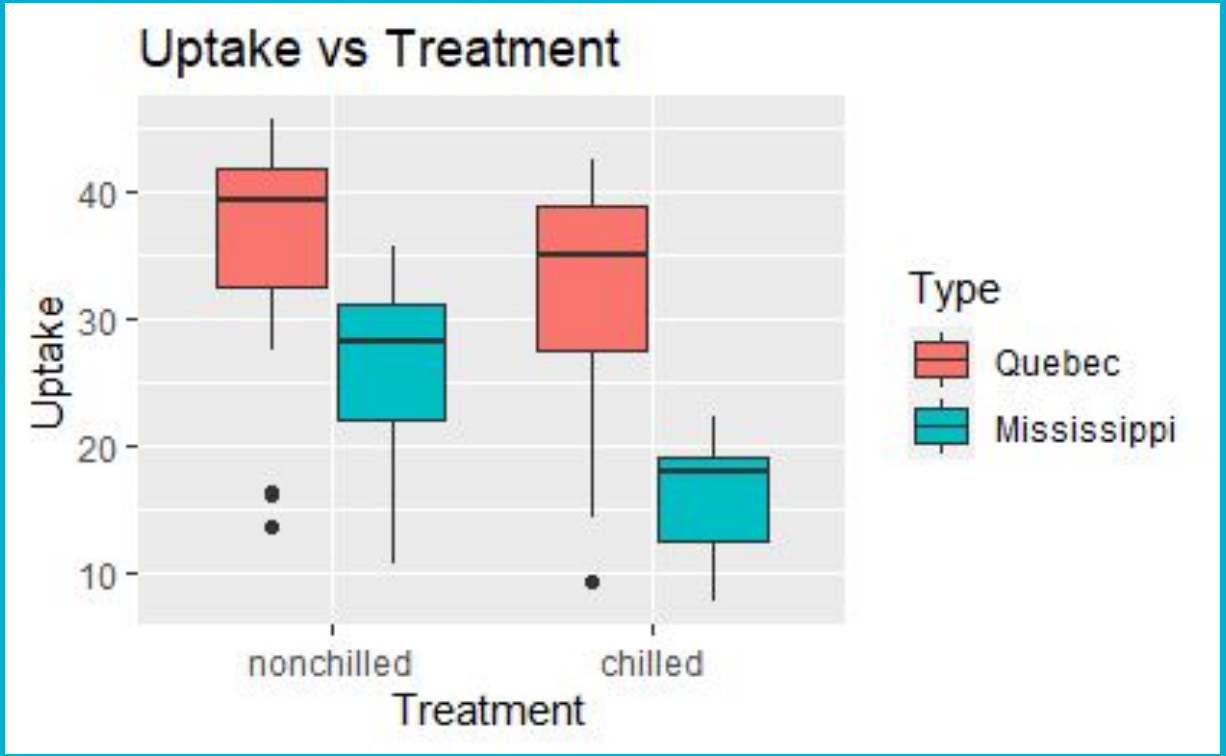


An important step in this bar graph would be to manually change the colors of the bar to determine the difference between the groups

The legend is also needed to show the readers which color corresponds with what group

geom_boxplot

From data set CO2



geom_boxplot code

Code:

```
library(lattice)

library(ggplot2)

library(tidyr)

data("CO2")

ggplot(CO2, aes(x = Treatment, y = uptake, fill=Type)) +
geom_boxplot()+xlab("Treatment")+ylab("Uptake")+ggtitle("Uptake vs Treatment")
```

geom_boxplot arguments

1. mapping: Aesthetic mappings
2. data: Data
3. position: Adjusts the position
4. outlier.color, outlier.fill, outlier.shape, outlier.size, outlier.stroke, outlier.alpha: Changes the aesthetics for outliers
5. notch: If set to true will make a notched box plot to compare groups
6. notchwidth: For a notched box plot sets the width of the body

Geom_boxplot arguments

7. varwidth: If true, set widths proportional to the square roots of the number of observations in the group

8. na.rm: Silently removes the missing values if set to true

9. orientation: Orientation of the layer

10. show.legend: Decide if layer should be included in the legends

11. inherit.aes: If False it will override the default aesthetics and not combine with them

12: coef: Length of the whiskers as multiple IQR. Default is 1.5

geom_violin

This is called a violin plot. It is essentially a mixture of the geom_boxplot and geom_density visualizations.

You can put a boxplot inside of the violin plot and specify the width.

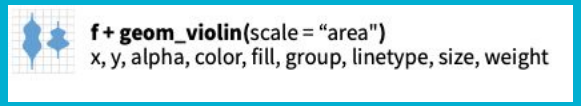
I put aesthetics inside of the violin plot to be colored as the different trees

```
{r}
data(Orange)

library(ggplot2)

Orange$age = as.numeric(Orange$age)
Orange$circumference = as.numeric(Orange$circumference)
Orange$Tree = factor(Orange$Tree, levels = c("1","2","3","4","5"))

p = ggplot(Orange, aes(Tree, circumference))+
  geom_violin(aes(fill = Tree))+
  geom_boxplot(width = .1)+
  labs(title = "Orange Trees")
p
```

Using levels, I ordered the trees from 1 to 5 on the X axis

With labs I put a title

More info on violin plots

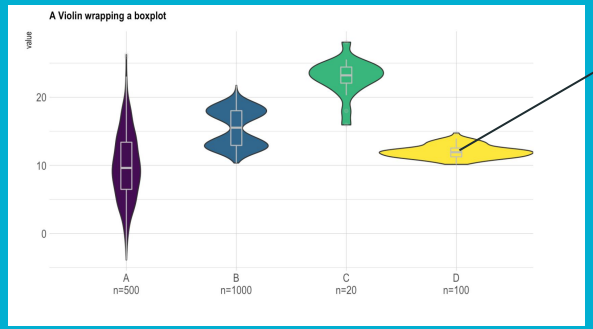


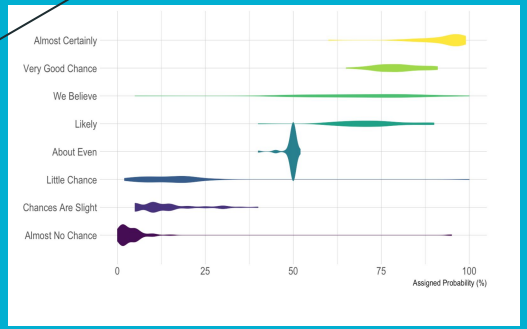
a categorical variable for the X axis: it needs to be have the class factor

- a numeric variable for the Y axis: it needs to have the class numeric

- should be in long format

Different examples of representing data with violin plots:





Thicker portion of a violin plot means higher frequency.

We could use a density plot, but when too many colors are overlaid, it is hard to read. Good for many categories.

With just a mean, we can't see the distribution of a data set; with a violin plot we can. This is a more accurate portrayal of the data.

How do we read this type of plot:

https://www.youtube.com/watch?v=M6Nu59Fsyyw&t=79s

Resources

- Bar graph: http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization
- Violin: https://ggplot2.tidyverse.org/reference/geom_violin.html
- Boxplot: http://www.sthda.com/english/wiki/ggplot2-box-plot-quick-start-guide-r-software-and-data-visualization