IntroBioCompExercise9\_Kodak

Haley Kodak

November 5, 2018

# Exercise 9 - Analysis and plotting

1. Find some data on two variables that you would expect to be related to each other. These data can come from your own research, your daily life, or the internet. Enter those data into a text ???le or Excel and then save a text ???le, and write a script that loads this text ???le and produces a scatter plot of those two variables that includes a trend line.

library(ggplot2)  
timber = read.csv("timber.csv", header = TRUE)  
str(timber)

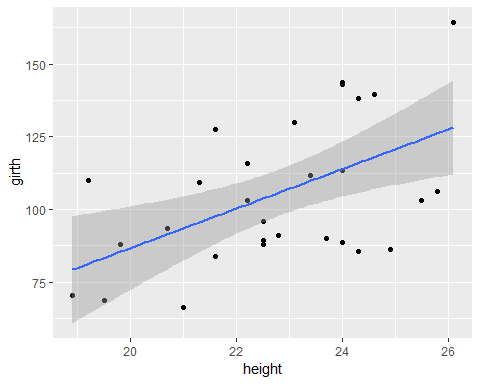
## 'data.frame': 31 obs. of 3 variables:  
## $ volume: num 0.746 0.746 0.739 1.188 1.361 ...  
## $ girth : num 66.2 68.6 70.2 83.8 85.4 ...  
## $ height: num 21 19.5 18.9 21.6 24.3 24.9 19.8 22.5 24 22.5 ...

head(timber)

## volume girth height  
## 1 0.7458 66.23 21.0  
## 2 0.7458 68.62 19.5  
## 3 0.7386 70.22 18.9  
## 4 1.1875 83.79 21.6  
## 5 1.3613 85.38 24.3  
## 6 1.4265 86.18 24.9

I will look at the relatinship between girth and hieght of trees

x=timber$height #assign x and y   
y=timber$girth  
  
a=ggplot(data=timber,aes(x=height,y=girth)) #assign labels and graph  
a+geom\_point()+coord\_cartesian()+stat\_smooth(method="lm",formula=y~x)



#plot data points and trendline

1. Given the data in “data.txt”. Write a script that generates two ???gures that sumamrize the data. First, show a barplot of the means of the four populations. Second, show a scatter plot of all of the observations. You may want to “jitter” the points (geom\_jitter()) to make it easier to see all of the observations within a population in your scatter plot. Alternatively, you could also try setting the alpha argument in geom\_scatterplot() to 0.1. Do the bar and scatter plots tell you di???erent stories? Why?

data = read.table("data.txt", sep = ",", header = TRUE)  
str(data)

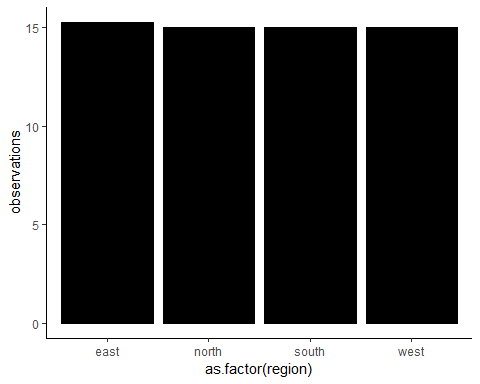
## 'data.frame': 4000 obs. of 2 variables:  
## $ region : Factor w/ 4 levels "east","north",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ observations: num 15.3 14.5 14.2 15.6 15.8 ...

head(data)

## region observations  
## 1 north 15.32186  
## 2 north 14.46967  
## 3 north 14.23646  
## 4 north 15.57369  
## 5 north 15.80087  
## 6 north 15.02925

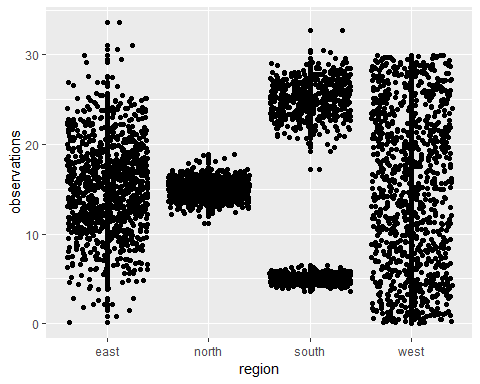
Barplot of means

meanObservations= mean(data$observations)  
b=d=ggplot(data=data)  
b+geom\_bar(aes(x=as.factor(region),y=observations),stat="summary",fun.y="mean",fill='black',color='black')+theme\_classic()



Scatter plot of all the observations

x=data$region #assign x and y   
y=data$observations  
  
c=ggplot(data=data,aes(x=region,y=observations)) #assign labels and graph  
c+geom\_point()+coord\_cartesian()+stat\_smooth(method="lm",formula=y~x)+geom\_jitter()



#plot data points and trendline

The bar plot allows comparison of the mean, but the scatter plot allows comparrison of the spread of the data for each region