9/11/2020

# libraries needed  
library(ggplot2)  
library(readxl)  
library(tidyverse)

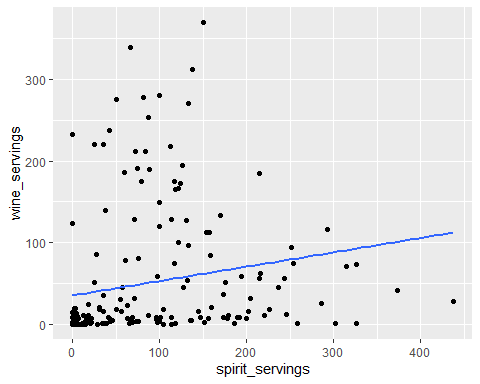
## -- Attaching packages ---------------------------------------------------- tidyverse 1.3.0 --

## v tibble 3.0.3 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0  
## v purrr 0.3.4

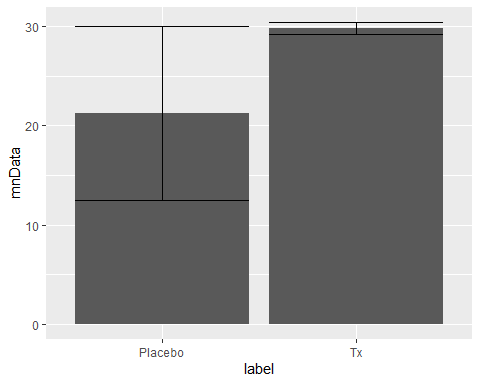
## -- Conflicts ------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

#1) Please make a scatterplot that compares any of the two variables in the drinks.csv data sets.  
dataFileName='./datastore/drinks.csv'  
drinks = read.csv(dataFileName,header=T)  
  
graph <- ggplot(drinks,aes(x=spirit\_servings, y=wine\_servings))   
graph + geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE)

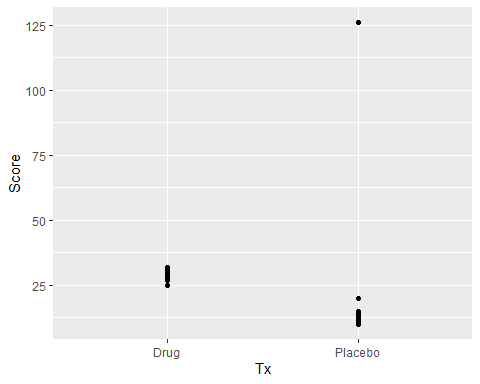
## `geom\_smooth()` using formula 'y ~ x'



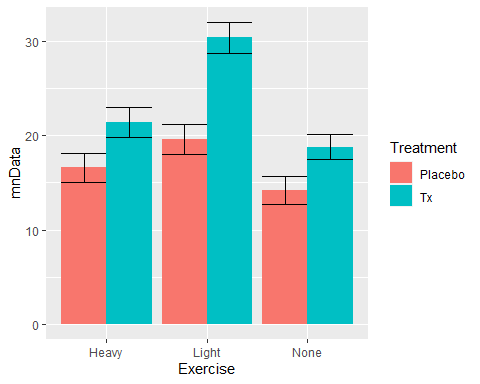
## `geom\_smooth()` using formula 'y ~ x'  
  
#2) Please open the Novel\_SSRI.xls data set and make a plot where the bars represent the group means and the errorbars are the standard error of the mean of each group.  
Novel\_SSRI <- read\_excel("./datastore/Novel\_SSRI.xls")  
## get into Tidy format  
Novel\_SSRI\_tidy <- gather(data=Novel\_SSRI, key = Tx, value = Score, -subjectNumber)  
# For Tidy format could also use pipe operator %>&=%  
Novel\_SSRI\_tidy <- Novel\_SSRI %>%  
 gather(key = Tx, value = Score, -subjectNumber)  
  
## Make Data Frame with a Mn and SEM for each group  
# create variable mnData which has the means  
mnData=c(mean(Novel\_SSRI\_tidy[Novel\_SSRI\_tidy$Tx=='Placebo',]$Score), +  
 mean(Novel\_SSRI\_tidy[Novel\_SSRI\_tidy$Tx=='Drug',]$Score))  
  
# create variable sdData which has the St. Dev.   
sdData=c(sd(Novel\_SSRI\_tidy[Novel\_SSRI\_tidy$Tx=='Placebo',]$Score), +  
 sd(Novel\_SSRI\_tidy[Novel\_SSRI\_tidy$Tx=='Drug',]$Score))  
  
# create variable semData which has the SEM   
semData=sdData/sqrt(13) # <-- 13 = Bad habit!!!  
  
# Create a label for each row  
label=c("Placebo","Tx")  
#Create a data frame with all of the needed info  
ssriData=data.frame(label,mnData,sdData,semData)  
  
## Plot the data  
graph <- ggplot(ssriData,aes(label,mnData))  
graph + geom\_bar(stat="identity") + # stat="identity" tells R to use the values provided instead of counts (default)  
 geom\_errorbar(aes(ymin = mnData-semData, ymax = mnData+semData))



## Notice the huge error bars in the Placebo group?   
# plot each data point  
graph <- ggplot(Novel\_SSRI\_tidy,aes(Tx,Score))  
graph + geom\_point(stat="identity")



## Indicates a problem. Remove the data and replot  
Novel\_SSRI\_tidy\_clean=Novel\_SSRI\_tidy[Novel\_SSRI\_tidy$subjectNumber!=11,]  
## repeat steps above to compute mean and sd  
  
#3) Please open the DeprScore.csv data set and make a plot where the bars represent the group means and the errorbars are the standard error of the mean of each group. Please plot the bars that refer to treatment in different colors.  
## make the data frame  
#means  
dataFileName='./datastore/DeprScore.csv'  
DeprScore = read.csv(dataFileName,header=T)  
mnData=c(mean(DeprScore[DeprScore$Exercise=='None' & DeprScore$Treatment=='Placebo',]$Score), +  
 mean(DeprScore[DeprScore$Exercise=='Light' & DeprScore$Treatment=='Placebo',]$Score), +  
 mean(DeprScore[DeprScore$Exercise=='Heavy' & DeprScore$Treatment=='Placebo',]$Score), +  
 mean(DeprScore[DeprScore$Exercise=='None' & DeprScore$Treatment=='Tx',]$Score), +  
 mean(DeprScore[DeprScore$Exercise=='Light' & DeprScore$Treatment=='Tx',]$Score), +  
 mean(DeprScore[DeprScore$Exercise=='Heavy' & DeprScore$Treatment=='Tx',]$Score))  
  
#sd  
sdData=c(sd(DeprScore[DeprScore$Exercise=='None' & DeprScore$Treatment=='Placebo',]$Score), +  
 sd(DeprScore[DeprScore$Exercise=='Light' & DeprScore$Treatment=='Placebo',]$Score), +  
 sd(DeprScore[DeprScore$Exercise=='Heavy' & DeprScore$Treatment=='Placebo',]$Score), +  
 sd(DeprScore[DeprScore$Exercise=='None' & DeprScore$Treatment=='Tx',]$Score), +  
 sd(DeprScore[DeprScore$Exercise=='Light' & DeprScore$Treatment=='Tx',]$Score), +  
 sd(DeprScore[DeprScore$Exercise=='Heavy' & DeprScore$Treatment=='Tx',]$Score))  
#sem  
semData=sdData/sqrt(5) #<-- bad habit!  
  
# Create a label for each row  
Exercise=c("None","Heavy","Light","None","Heavy","Light")  
Treatment=c("Placebo","Placebo","Placebo","Tx","Tx","Tx")  
  
#Create a data frame with all of the needed info  
dScore=data.frame(Exercise,Treatment,mnData,sdData,semData)  
  
#Plot the data  
ggplot(data = dScore, aes(x = Exercise, y = mnData, fill = Treatment)) +  
 geom\_bar(stat='identity',position = "dodge") +  
 geom\_errorbar(aes(ymin = mnData-semData, ymax = mnData+semData), position = "dodge")



## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.