STAC32 Assignment #1

Grisham Nathan 1001314927

September 14, 2017

# a)

library(tidyverse)

## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: readr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats  
## lag(): dplyr, stats

url="http://www.utsc.utoronto.ca/~butler/c32/global.csv"  
input=read\_csv(url)

## Parsed with column specification:  
## cols(  
## warehouse = col\_character(),  
## size = col\_integer(),  
## cost = col\_double()  
## )

input

## # A tibble: 10 x 3  
## warehouse size cost  
## <chr> <int> <dbl>  
## 1 A 225 11.95  
## 2 B 350 14.13  
## 3 A 150 8.93  
## 4 A 200 10.98  
## 5 A 175 10.03  
## 6 A 180 10.13  
## 7 B 325 13.75  
## 8 B 290 13.30  
## 9 B 400 15.00  
## 10 A 125 7.97

The data frame has 10 rows and 3 columns.

# b)

ggplot(input,aes(x=size,y=cost))+geom\_point()



# c)

There is a positive relationship between size and cost. As size increases, cost also increases.

The trend appears to go from bottom left to top right.  
A straight regression line would describe the trend appropriately, however their are some points that cause variance from the regression line.

# d)

ggplot(input,aes(x=size,y=cost,colour=warehouse))+geom\_point()

I think size determines which warehouse an incoming shipment goes to. From the data given and the scatterplot I created, the largest shipment to warehouse A is 225000 parts and the smallest shipment to warehouse B is 290000 parts. It seems that whether a shipment has more or less than a certain number of parts (Definitely between 225000 and 290000 parts), is what determines which warehouse the shipment goes. As a result, we can infer that products below this specified limit goes to warehouse A and products above this specified limit go to warehouse B.