EDA Casino Data

Aakanksha Baid

1/27/2020

## Step 1a: Import the data

# import  
casino\_DF <- read.csv("Data/casino.csv")

## Step 1b: Install the package tidyverse

#install.packages("tidyverse")  
library(tidyverse)

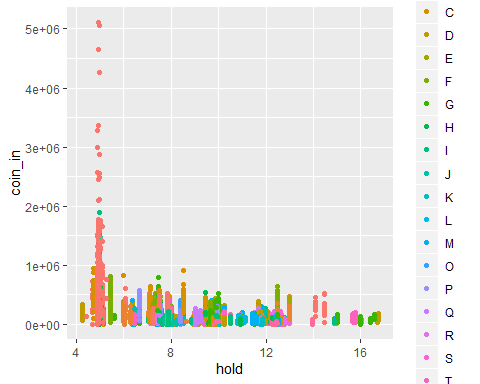
## -- Attaching packages ------------------------------------------------------------ tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

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

## Step 2: Scatter plot 1

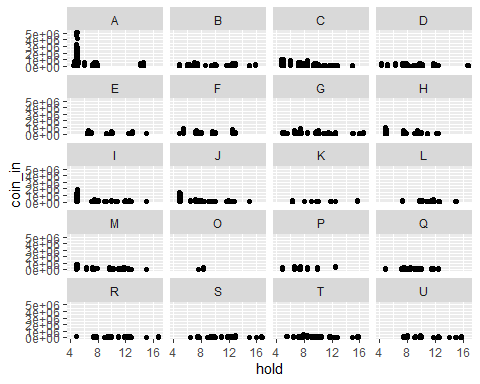
# hold vs coin\_in colored by location  
ggplot(data=casino\_DF) +  
 geom\_point(mapping = aes(x=hold,y=coin\_in,color=location))



The above plot shows a negative relationship between hold and coin-ins.

## Step 3: Scatter plot 2

# hold vs coin\_in for each location  
ggplot(data=casino\_DF) +  
 geom\_point(mapping = aes(x=hold,y=coin\_in)) +  
 facet\_wrap(~ location,nrow = 5)



The above plot shows that the relationship between hold and coin-in is in fact not negatively related for all locations, when we factor in location.

## Step 4: Interpret the plots

1. The relationship between holds and coin-in of slot machines varies across locations as location A has high coin-in even at low hold which is also the case for locations I and J, whereas for other locations, hold is fairly constant at low coin-ins. This is not clear from Step 2 plot.
2. The Step 3 plot is better to interpret relationships between the variables as it clearly displays the relationship between coin-in and hold for each location and the graph is intuitive. The step 2 plot is not visually appealing as it is not able to display the large number of locations on the same plot (eg. location A is missing) making it difficult to identify any meaningful relationship due to high number of locations.