initial\_data\_exploration.R

Yash Tambawala

Fri Feb 24 14:23:57 2017

# This file will do an initial exploratory analysis of the dataset to uncover basic insights   
# into the SBA loan datasets.   
  
# Authors: Daniel, Brent, Peng and Yash  
  
# File Description:   
# Inputs:   
  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(GGally)

## Warning: package 'GGally' was built under R version 3.3.2

##   
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':  
##   
## nasa

library(survival)  
library(ROCR)

## Warning: package 'ROCR' was built under R version 3.3.2

## Loading required package: gplots

## Warning: package 'gplots' was built under R version 3.3.2

##   
## Attaching package: 'gplots'

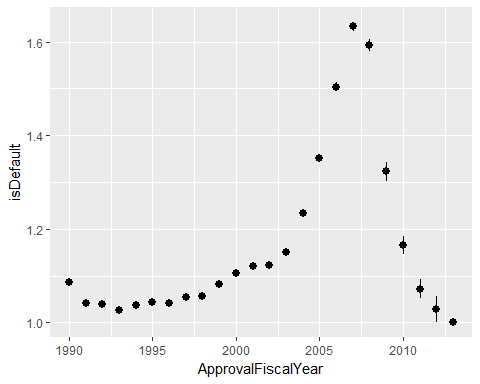
## The following object is masked from 'package:stats':  
##   
## lowess

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2

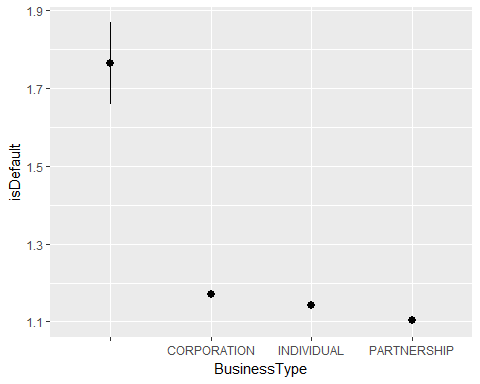
set.seed(1)  
  
df <- read.csv("SBA\_cleaned\_data.csv")  
  
ggplot(data = df)+stat\_summary(mapping = aes(x = ApprovalFiscalYear, y = isDefault))

## No summary function supplied, defaulting to `mean\_se()



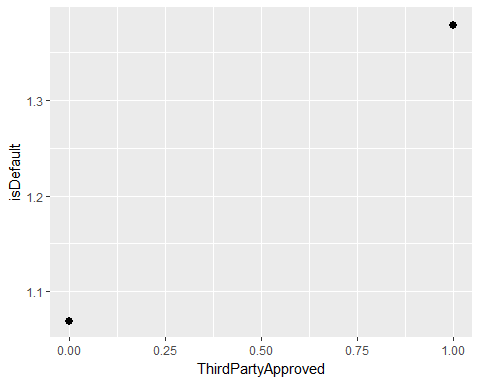
ggplot(data = df)+stat\_summary(mapping = aes(x = BusinessType, y = isDefault))

## No summary function supplied, defaulting to `mean\_se()



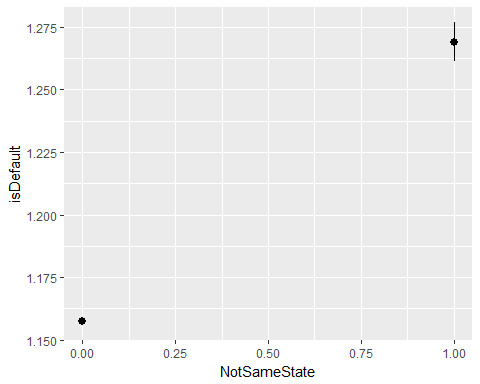
ggplot(data = df)+stat\_summary(mapping = aes(x = ThirdPartyApproved, y = isDefault))

## No summary function supplied, defaulting to `mean\_se()

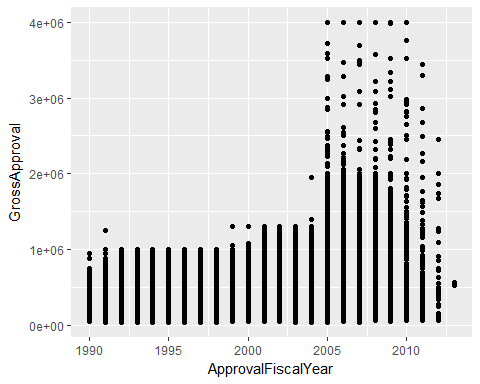


ggplot(data = df)+stat\_summary(mapping = aes(x = NotSameState, y = isDefault))

## No summary function supplied, defaulting to `mean\_se()



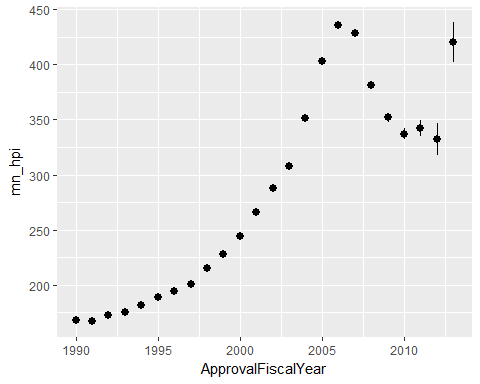
ggplot(data = df)+geom\_point(mapping = aes(x = ApprovalFiscalYear, y = GrossApproval))



ggplot(data = df)+stat\_summary(mapping = aes(x = ApprovalFiscalYear, y = mn\_hpi))

## Warning: Removed 360 rows containing non-finite values (stat\_summary).

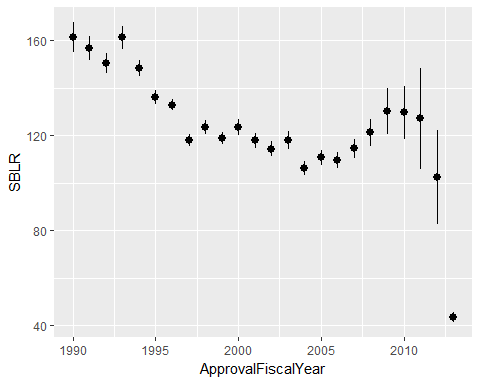
## No summary function supplied, defaulting to `mean\_se()



ggplot(data = df)+stat\_summary(mapping = aes(x = ApprovalFiscalYear, y = SBLR))

## Warning: Removed 359 rows containing non-finite values (stat\_summary).

## No summary function supplied, defaulting to `mean\_se()



ggplot(data = df)+stat\_summary(mapping = aes(x = Naics2digits, y = isDefault))

## Warning: Removed 16850 rows containing non-finite values (stat\_summary).

## No summary function supplied, defaulting to `mean\_se()

