auto-insurance

Krista Waugh

3/7/2019

library(readr)  
library(ggplot2)  
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(tidyverse)

## ── Attaching packages ───────────────────────────────────────────────────────────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ tibble 1.4.2 ✔ purrr 0.2.5  
## ✔ tidyr 0.8.1 ✔ stringr 1.3.1  
## ✔ tibble 1.4.2 ✔ forcats 0.3.0

## ── Conflicts ──────────────────────────────────────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

data <- read\_csv("decal\_midterm.csv")

## Parsed with column specification:  
## cols(  
## `Claim Number` = col\_double(),  
## AGE = col\_double(),  
## INCOME = col\_character(),  
## MARITAL\_STATUS = col\_character(),  
## EDUCATION = col\_character(),  
## MVR\_POINTS = col\_double(),  
## CAR\_AGE = col\_double(),  
## LOCATION = col\_character(),  
## OLD\_CLAIM = col\_character(),  
## PREMIUM = col\_character()  
## )

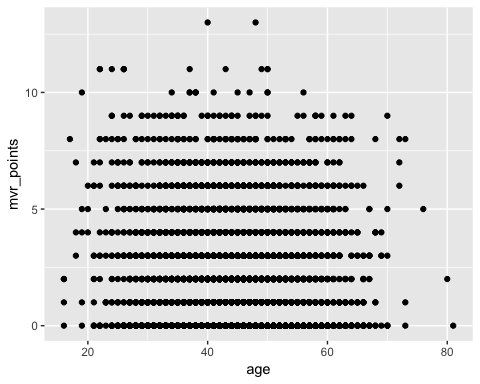
colnames(data) = c("claim\_number", "age", "income", "marital\_status", "education", "mvr\_points", "car\_age", "location", "old\_claim", "premium")  
data

## # A tibble: 8,161 x 10  
## claim\_number age income marital\_status education mvr\_points car\_age  
## <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>  
## 1 1 60 $67,3… No PhD 3 18  
## 2 2 43 $91,4… No High Sch… 0 1  
## 3 3 35 $16,0… Yes High Sch… 3 10  
## 4 4 51 $64,6… Yes <High Sc… 0 6  
## 5 5 50 $114,… Yes PhD 3 17  
## 6 6 34 $125,… No Bachelors 0 7  
## 7 7 54 $18,7… Yes <High Sc… 0 1  
## 8 8 37 $107,… Yes Bachelors 10 7  
## 9 9 34 $62,9… No Bachelors 0 1  
## 10 10 50 $106,… No Bachelors 1 17  
## # ... with 8,151 more rows, and 3 more variables: location <chr>,  
## # old\_claim <chr>, premium <chr>

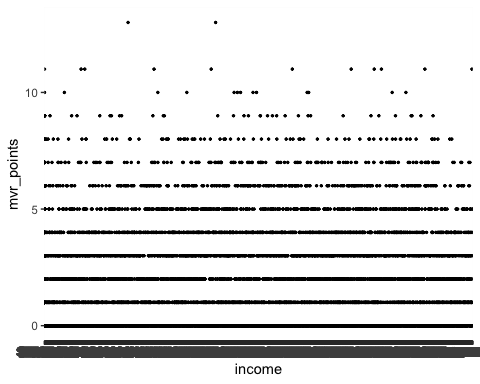
age <- select(data, age, mvr\_points)  
age

## # A tibble: 8,161 x 2  
## age mvr\_points  
## <dbl> <dbl>  
## 1 60 3  
## 2 43 0  
## 3 35 3  
## 4 51 0  
## 5 50 3  
## 6 34 0  
## 7 54 0  
## 8 37 10  
## 9 34 0  
## 10 50 1  
## # ... with 8,151 more rows

ggplot(age, aes(age, mvr\_points)) + geom\_point()

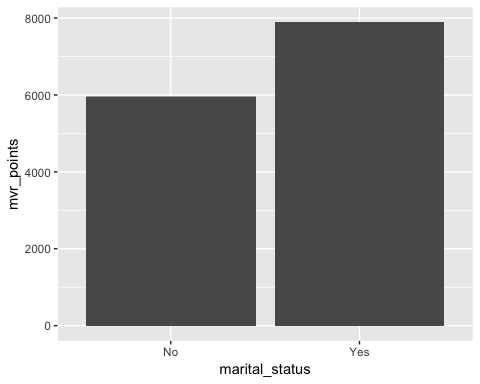


income <- select(data, income, mvr\_points)  
  
ggplot(income, aes(income, mvr\_points)) + geom\_point(size = .5)



marital\_status <- select(data, marital\_status, mvr\_points)  
  
ggplot(marital\_status, aes(marital\_status, mvr\_points)) + geom\_col(binwidth = 2)

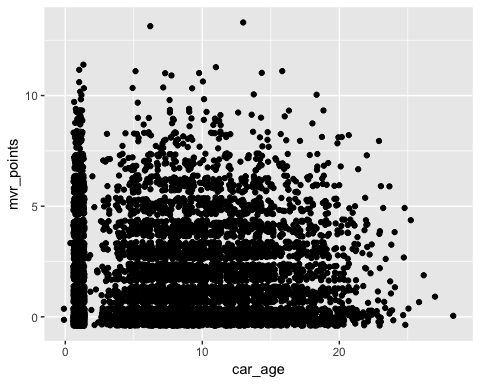
## Warning: Ignoring unknown parameters: binwidth



phd <- filter(data, education == "PhD")  
highschool <- filter(data, education == "High School")  
bachelor <- filter(data, education == "Bachelors")  
under\_highschool <- filter(data, education == "<High School")

education <-   
 ggplot(, aes(y = data$mvr\_points)) +  
 geom\_col(x = phd) +  
 geom\_col(x = bachelor) +  
 geom\_col(x = highschool) +  
 geom\_col(x = under\_highschool)  
education

car <- select(data, car\_age, mvr\_points)  
  
ggplot(car, aes(car\_age, mvr\_points)) + geom\_jitter()



rural\_urban <-   
 select(data, )

phd <- filter(data, education == "PhD")  
highschool <- filter(data, education == "High School")  
bachelor <- filter(data, education == "Bachelors")  
under\_highschool <- filter(data, education == "<High School")