

R Notebook

[Code ▾](#)

This is an [R Markdown](#) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

[Hide](#)

```
library(dplyr)
library(stringr)
#import data
montana <- read.csv("MT_cleaned.csv", stringsAsFactors = FALSE)
montanaT <- montana
vermont <- read.csv("VT_cleaned.csv", stringsAsFactors = FALSE)
vermontT <- vermont
```

[Hide](#)

```
#understand data
dim(montana)
str(montana)
head(montana)
colnames(montana)
```

[Hide](#)

```
# proportion of male drivers stop in MT
prop_m_stop = sum(montana$driver_gender=='M') / dim(montana)[1]
print(prop_m_stop, digits = 10)
```

```
[1] 0.6749749733
```

[Hide](#)

```
# arresting comparison between non_MT plate and MT plate
m <- subset(montana, out_of_state=='TRUE' & montana$is_arrested=='TRUE')
n <- subset(montana, out_of_state=='FALSE' & montana$is_arrested=='TRUE')
non_MT_arrt = dim(m)[1] / dim(n)[1]
print(non_MT_arrt, digits = 10)
```

[Hide](#)

```
# chi test for non_MT and MT arresting
chisq.test(table(montana$out_of_state=='TRUE' & montana$is_arrested=='TRUE',
                 montana$out_of_state=='FALSE' & montana$is_arrested=='TRUE'))
```

[Hide](#)

```
# proportion of speeding
prop_speeding <- sum(montana$violation=='Speeding') / dim(montana)[1]
print(prop_speeding, digits = 10)
```

[Hide](#)

```
# proportion of DUI in VT
prop_DUI_vt <- sum(vermont$violation %in% c('DUI'))/dim(vermont)[1]
print(prop_DUI_vt, digits = 10)
```

Hide

```
# linear regression model between year and average_manufacture_vehicle
# amv stands for average_manufacture_vehicle
# extract year from date
montana$year_stop <- format(as.Date(montana$stop_date), format="%Y")
table(montana$year_stop)

montana$year_cars <- as.numeric((montana$vehicle_year))

amv_09 <-
  round(mean(montana$year_cars[which(montana$year_stop=='2009')], na.rm = TRUE), 0)
amv_10 <-
  round(mean(montana$year_cars[which(montana$year_stop=='2010')], na.rm = TRUE), 0)
amv_11 <-
  round(mean(montana$year_cars[which(montana$year_stop=='2011')], na.rm = TRUE), 0)
amv_12 <-
  round(mean(montana$year_cars[which(montana$year_stop=='2012')], na.rm = TRUE), 0)
amv_13 <-
  round(mean(montana$year_cars[which(montana$year_stop=='2013')], na.rm = TRUE), 0)
amv_14 <-
  round(mean(montana$year_cars[which(montana$year_stop=='2014')], na.rm = TRUE), 0)
amv_15 <-
  round(mean(montana$year_cars[which(montana$year_stop=='2015')], na.rm = TRUE), 0)
amv_16 <-
  round(mean(montana$year_cars[which(montana$year_stop=='2016')], na.rm = TRUE), 0)

Year <-c('2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016')
average_manufacture_vehicle <- c(amv_09, amv_10, amv_11, amv_12, amv_13, amv_14, amv_15, amv_16)
dataT <- data.frame(Year, average_manufacture_vehicle)
View(dataT)

mod1 <- lm(Year ~ average_manufacture_vehicle)
summary(mod1)
# make prediction with year as variable
avm_20 <- (2020 + 860.26415)/1.43396
print(avm_20, digits = 10)
```

Hide

```
# import the combined data by operate cmd
# understand the combined data
data_comb <- read.csv("MT_VT_combine.csv", stringsAsFactors = FALSE)
dim(data_comb)
head(data_comb)
View(data_comb)[1:20]
str(data_comb)

# extract hours from the combined data
Split <- strsplit(as.character(data_comb$stop_time), ":", fixed = TRUE)
data_comb$stop_hs <- sapply(Split, "[", 1)
table(data_comb$stop_hs)

Split <- strsplit(as.character(montana$stop_time), ":", fixed = TRUE)
montana$stop_hs <- sapply(Split, "[", 1)
sort(table(montana$stop_hs))

diff_stop_num=75707 -229
print(diff_stop_num, digits = 10)
```

Hide

```
#predict county area with longitude and latitude
data5 <- group_by(montana, county_name)
head(data5)
new_f <-summarise(data5,
                  count=n(),
                  lat_sd=sd(lat, na.rm = TRUE),
                  lon_sd=sd(lon, na.rm = TRUE))
View(new_f)
new_f$size_sqkm <- 3.14 * 2 *new_f$lat_sd * 2 * new_f$lon_sd *10
max(new_f$size_sqkm)
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.