In class assignment 3

Diane Nguyen

August 31, 2018

```
knitr::opts_chunk$set(echo = TRUE)
#install.packages("tidyverse")
#install.packages("devtools")
library(devtools)
#install_github("uc-cfss/rcfss")
#help(?gun_deaths)
library(rcfss)
data("gun_deaths")
str(gun_deaths)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    100798 obs. of  10 variables:
##  $ id       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ year     : int  2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...
##  $ month    : num  1 1 1 2 2 2 2 3 2 2 ...
##  $ intent   : chr  "Suicide" "Suicide" "Suicide" "Suicide" ...
##  $ police   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ sex      : chr  "M" "F" "M" "M" ...
##  $ age      : int  34 21 60 64 31 17 48 41 50 NA ...
##  $ race     : chr  "Asian/Pacific Islander" "White" "White" "White" ...
##  $ place    : chr  "Home" "Street" "Other specified" "Home" ...
##  $ education: Factor w/ 4 levels "Less than HS",..: 4 3 4 4 2 1 2 2 3 NA ...
```

```
colnames(gun_deaths)
```

```
##  [1] "id"        "year"      "month"     "intent"    "police"
##  [6] "sex"       "age"       "race"      "place"     "education"
```

```
library(tidyverse)#install to make tables look pretty
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(plyr)
```

```
## -------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following object is masked from 'package:purrr':
##
##     compact
```

```
library(dplyr)
library(tidyr)
library(tibble)
library(ggplot2)#helps with graphing stuff
library(knitr) #this one gets data about the layout of your info
as.tibble(gun_deaths)
```

```
## # A tibble: 100,798 x 10
##       id  year month intent  police sex     age race      place  education
##    <int> <int> <dbl> <chr>    <int> <chr> <int> <chr>     <chr>  <fct>
##  1     1  2012     1 Suicide      0 M        34 Asian/Pa~ Home   BA+
##  2     2  2012     1 Suicide      0 F        21 White     Street Some col~
##  3     3  2012     1 Suicide      0 M        60 White     Other~ BA+
##  4     4  2012     2 Suicide      0 M        64 White     Home   BA+
##  5     5  2012     2 Suicide      0 M        31 White     Other~ HS/GED
##  6     6  2012     2 Suicide      0 M        17 Native A~ Home   Less tha~
##  7     7  2012     2 Undete~      0 M        48 White     Home   HS/GED
##  8     8  2012     3 Suicide      0 M        41 Native A~ Home   HS/GED
##  9     9  2012     2 Accide~      0 M        50 White     Other~ Some col~
## 10    10  2012     2 Suicide      0 M        NA Black     Home   <NA>
## # ... with 100,788 more rows
```

```
# must install all of these libraries to get the data to display correctly in the graphs
#must run libraries every time you open R studio for code to run properly
```

1. In what month do the most gun deaths occur?

```
knitr::opts_chunk$set(echo = TRUE)
#change your months to print out the name of the month
gun_deaths_month<-mutate(gun_deaths,
                         month_name=as.factor(mapvalues(month, c(1,2,3,4,5,6,7,8,9,10,11,12), c("Jan.","Fe
                                                                                                "Aug.","Sept."
gun_deaths_month_sum =dplyr::count(gun_deaths_month,month_name)

arrange(gun_deaths_month_sum,desc(n))%>%
  kable(caption="Gun deaths by month by descending order",
        col.names=c("Month","Number of deaths"))
```

Table 1: Gun deaths by month by descending order

Month	Number of deaths
July	8989
Aug.	8783
June	8677
May	8669
Sept.	8508
Apr.	8455
Dec.	8413
Oct.	8406
March	8289
Jan.	8273
Nov.	8243
Feb.	7093

2. What is the most common intent in gun deaths? Do most people killed by guns die in suicides, homicides, or accidental shootings?

```
knitr::opts_chunk$set(echo = TRUE)
intent_gun_death=dplyr::count(gun_deaths,intent)
arrange(intent_gun_death,desc(n))%>%
  kable(caption="Most common intent in gun deaths",
        col.names=c("intent","Number of deaths"))
```

Table 2: Most common intent in gun deaths

intent	Number of deaths
Suicide	63175
Homicide	35176
Accidental	1639
Undetermined	807
NA	1

3 . What is the average age of females killed by guns?

```
knitr::opts_chunk$set(echo = TRUE)
gun_data_women<-gun_deaths[gun_deaths$sex=="F", ]
mean(gun_data_women$age,na.rm=TRUE)%>%
  kable(caption="Avg. Age of females killed by guns",
        col.names="Age")
```

Table 3: Avg. Age of females killed by guns

Age
43.69507

4 How many white males with at least a high school education were killed by guns in 2012?

```
knitr::opts_chunk$set(echo = TRUE)
ed_male_gun<-subset(gun_deaths,education !="Less that high school"& sex=="M"& year=="2012" & race=="Whi
unique(ed_male_gun$education)
```

```
## [1] BA+          HS/GED       Some college Less than HS
## Levels: Less than HS HS/GED Some college BA+
```

```
gun_death_race=dplyr::count(ed_male_gun, race)
arrange(gun_death_race,desc(n))%>%
  kable(caption="White Males with at least high school education killed by guns in 2012",
      col.names=c("Race","Number of deaths"))
```

Table 4: White Males with at least high school education killed by guns in 2012

Race	Number of deaths
White	18057

5 Which season of the year has the most gun deaths? Assume that #. Winter = January-March #. Spring = April-June #. Summer = July-September #. Fall = October-December Hint: you need to convert a continuous variable into a categorical variable. Find a function that does that.

```
knitr::opts_chunk$set(echo = TRUE)
season_deaths<-mutate(gun_deaths,
                      season=as.factor(mapvalues(month, c(1,2,3,4,5,6,7,8,9,10,11,12), c("Winter","Winte
                                                 "Summer","Summer","Fall","Fall","Fall"))))
#mutate changes the value of the months to the season
season_deaths<-as.tibble(season_deaths)
season_gun_death=dplyr::count(season_deaths,season)
arrange(season_gun_death,desc(n))%>%
  kable(caption="Season of the year with the most gun deaths",
        col.names=c("Month","Number of deaths"))
```

Table 5: Season of the year with the most gun deaths

Month	Number of deaths
Summer	26280
Spring	25801
Fall	25062
Winter	23655

```
#mutate changes the value of the months to the season
```

6 What is the relationship between race and intent? For example, are whites who are killed by guns more likely to die because of suicide or homicide? How does this compare to blacks and hispanics?

```
knitr::opts_chunk$set(echo = TRUE)
table(gun_deaths$race,gun_deaths$intent, useNA="no")
```

```
##
##                                 Accidental Homicide Suicide Undetermined
##   Asian/Pacific Islander                12      559     745           10
##   Black                                328    19510    3332          126
##   Hispanic                             145     5634    3171           72
##   Native American/Native Alaskan        22      326     555           14
##   White                               1132     9147   55372          585
```

```
race_intent<-round(prop.table(table(gun_deaths$race,gun_deaths$intent, useNA="no"),1), digits=2)

race_intent<-as.data.frame.matrix(race_intent)
  race_intent%>% #gives you the result of gun deaths by race and intent
  kable(caption="Relationship between race and intent"
        )
```

Table 6: Relationship between race and intent

	Accidental	Homicide	Suicide	Undetermined
Asian/Pacific Islander	0.01	0.42	0.56	0.01
Black	0.01	0.84	0.14	0.01
Hispanic	0.02	0.62	0.35	0.01
Native American/Native Alaskan	0.02	0.36	0.61	0.02
White	0.02	0.14	0.84	0.01

7 Are police-involved gun deaths significantly different from other gun deaths? Assess the relationship between police involvement and age, police involvement and race, and the intersection of all three variables.

```
knitr::opts_chunk$set(echo = TRUE)
#have to map police involvment for this function to work. 0=none, 1=involved
#use the mutate function to accomplish this
police_deaths<-mutate(gun_deaths,
                      police=as.factor(mapvalues(police, c(0,1), c("Police not involved","police involve
police_deaths
```

```
## # A tibble: 100,798 x 10
##       id  year month intent  police   sex     age race    place  education
##    <int> <int> <dbl> <chr>   <fct>    <chr> <int> <chr>   <chr>  <fct>
##  1     1  2012     1 Suicide Police ~ M        34 Asian/~ Home   BA+
##  2     2  2012     1 Suicide Police ~ F        21 White   Street Some col~
##  3     3  2012     1 Suicide Police ~ M        60 White   Other~ BA+
##  4     4  2012     2 Suicide Police ~ M        64 White   Home   BA+
##  5     5  2012     2 Suicide Police ~ M        31 White   Other~ HS/GED
##  6     6  2012     2 Suicide Police ~ M        17 Native~ Home   Less tha~
##  7     7  2012     2 Undete~ Police ~ M        48 White   Home   HS/GED
##  8     8  2012     3 Suicide Police ~ M        41 Native~ Home   HS/GED
##  9     9  2012     2 Accide~ Police ~ M        50 White   Other~ Some col~
## 10    10  2012     2 Suicide Police ~ M        NA Black   Home   <NA>
## # ... with 100,788 more rows
```

```
police_three_data <- ddply(police_deaths, ~race+police, summarize,
                           mean.age= mean(age, na.rm = TRUE))

police_three_data%>%
  kable(caption="Police involvement by intent"
  )
```

Table 7: Police involvement by intent

race	police	mean.age
Asian/Pacific Islander	police involved	31.26667
Asian/Pacific Islander	Police not involved	38.78781

race	police	mean.age
Black	police involved	31.73034
Black	Police not involved	31.08789
Hispanic	police involved	33.34043
Hispanic	Police not involved	33.21238
Native American/Native Alaskan	police involved	30.60000
Native American/Native Alaskan	Police not involved	36.39169
White	police involved	39.69394
White	Police not involved	50.11715