Finial Project

Aspen Fairley

Coded libraries:

```r
library(dplyr)
library(readxl)
library(readr)
library(gridExtra)
library(chron)
library(ggplot2)
library(forcats)
```

Intro

I'm going to be looking at the police data for all the traffic stops in Connecticut during October 1, 2013 – May 31, 2014. Officers are required by law to take down this data when doing a stop of any kind. The dataset contains **366060** observations and **29** variables but I'm only going to focus on **3** of them: The age of divers stopped, whether or not the car had contaband and what happend to the drivers.

Reading in the Data

```r
load("/Users/aspen/Documents/math130/math315/Trafficstop_pds_Clean4.Rdata")
```

Univariate descriptions

Subject Age

```r
summary(Traffic_clean4$Subject_Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   12.00   26.00   36.00   38.33   49.00  100.00    1151
```

This is a table of summary statistics comparing drivers ages. From our table we can easily see that our age range keeps to the young side. Our median age is **36** years old with a min of **12** years and a max of **100** years. Our mean is almost surprisingly young at **38** years and most of our data is between **26** years old and **49** years. This can be seen even clearer with my graph.

```r
ggplot(Traffic_clean4, aes(x=Subject_Age)) + geom_histogram() + theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1151 rows containing non-finite values (stat_bin).
```
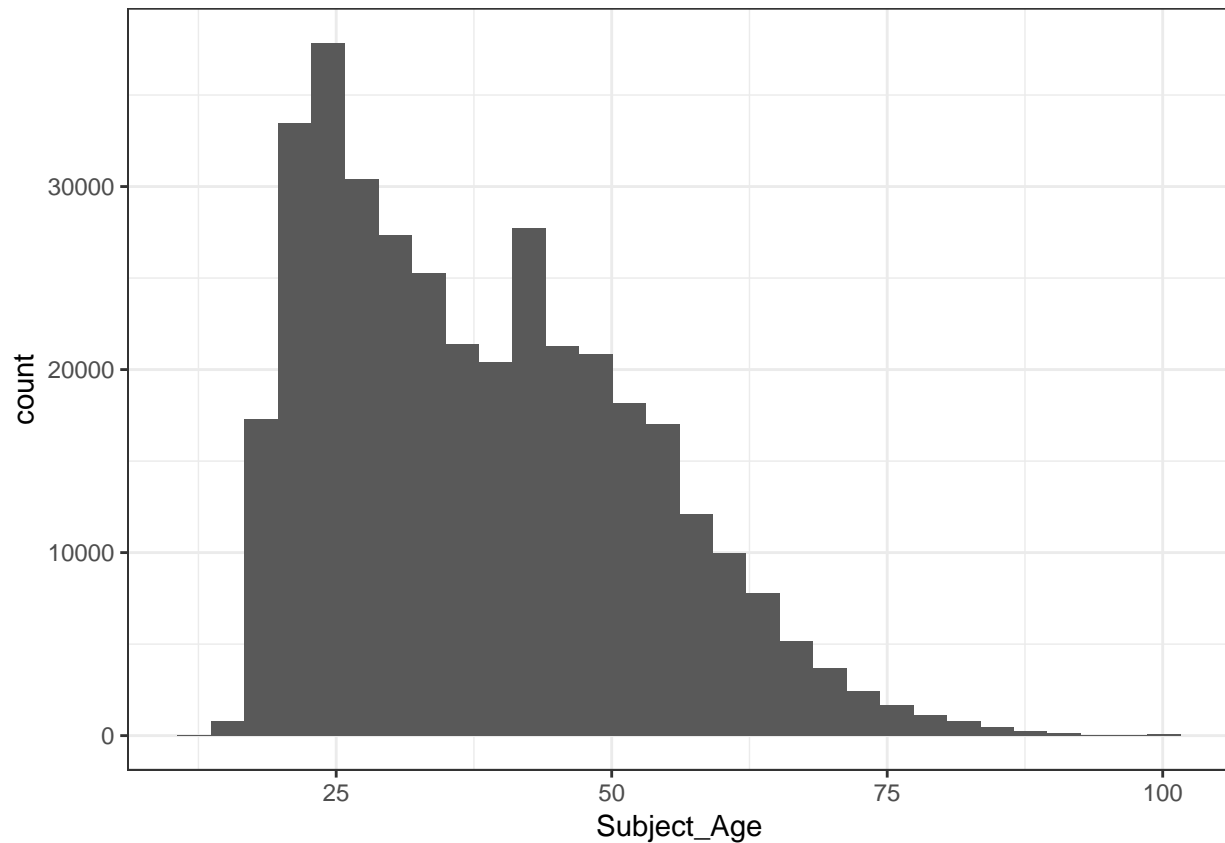
Here we can see my graph is skewed to the right.

Contraband

```
table(Traffic_clean4$Contraband)
```

```
##
##      0      1
## 361808   4248
```

As my contraband data is in binary, it will be hard to graph. I decided to change its class from integer to factor.

```
Traffic_clean4$ContrabandYN<- factor(Traffic_clean4$Contraband, labels=c("no", "yes"))
table(Traffic_clean4$ContrabandYN, Traffic_clean4$Contraband, useNA = "always")
```

```
##
##              0      1   <NA>
##   no     361808      0      0
##   yes         0   4248      0
##   <NA>        0      0      4
```

Now I can graph it.

```r
ggplot(Traffic_clean4, aes(x=ContrabandYN)) + geom_bar() + xlab("Contraband")
```
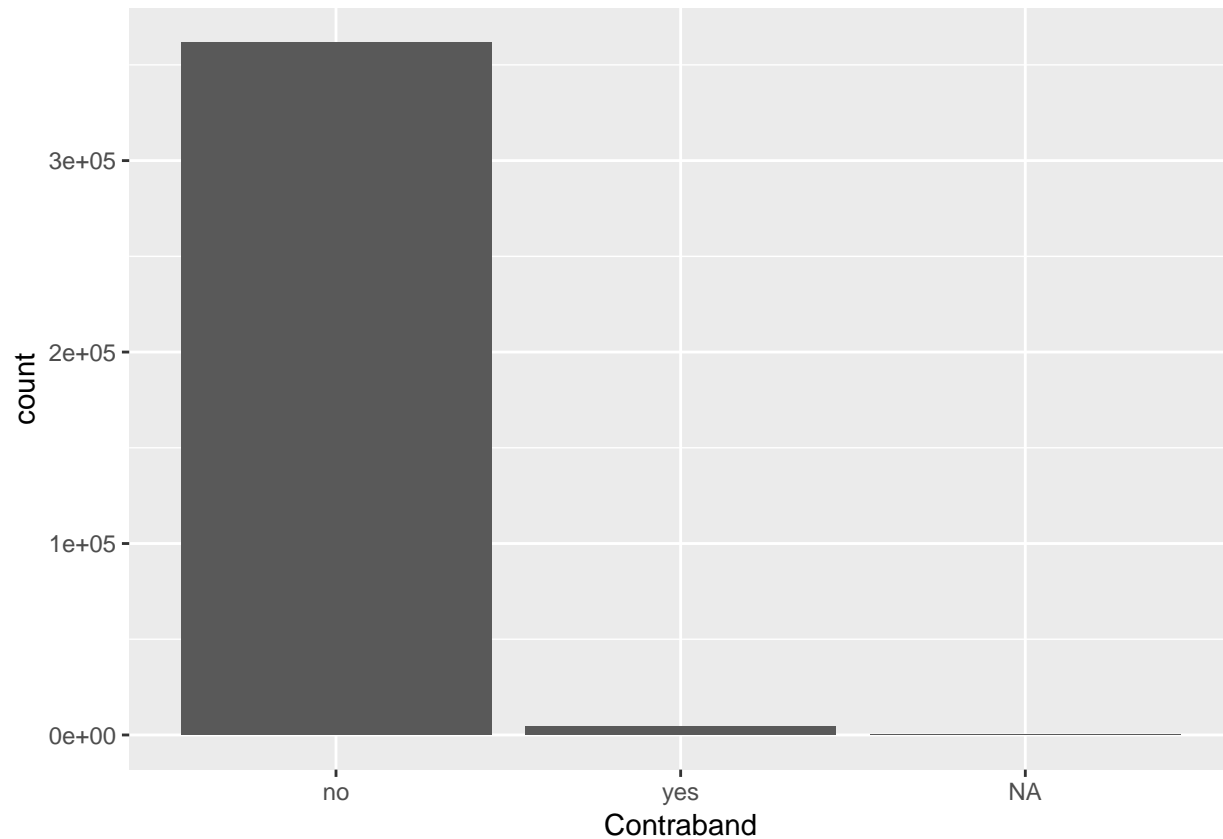

This table and graph is comparing the number of people who were caught to be carrying contraband to the number of people who were not. We can definitly see that, comparably, few people were found to have contraband.

Action

```r
Traffic_clean4$Dispostion_Cat %>% fct_infreq() %>% table()
```

```
##  .
##      ticket  verbal_warn written_warn      summons    none_done
##      170792        98489        66285        21041         6112
##      arrest
##        3311
```

```r
ggplot(Traffic_clean4, aes(x=forcats::fct_infreq(Dispostion_Cat))) + geom_bar() +
  xlab("Action")
```
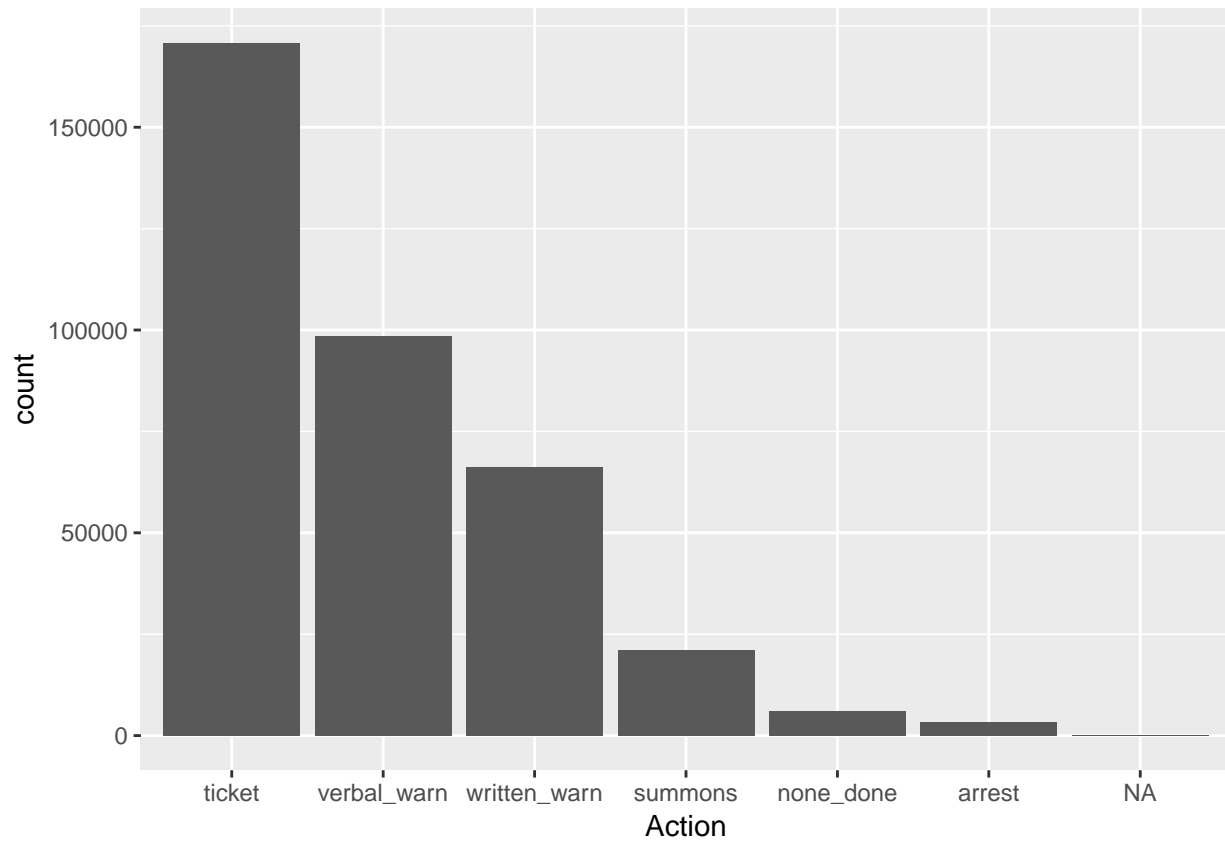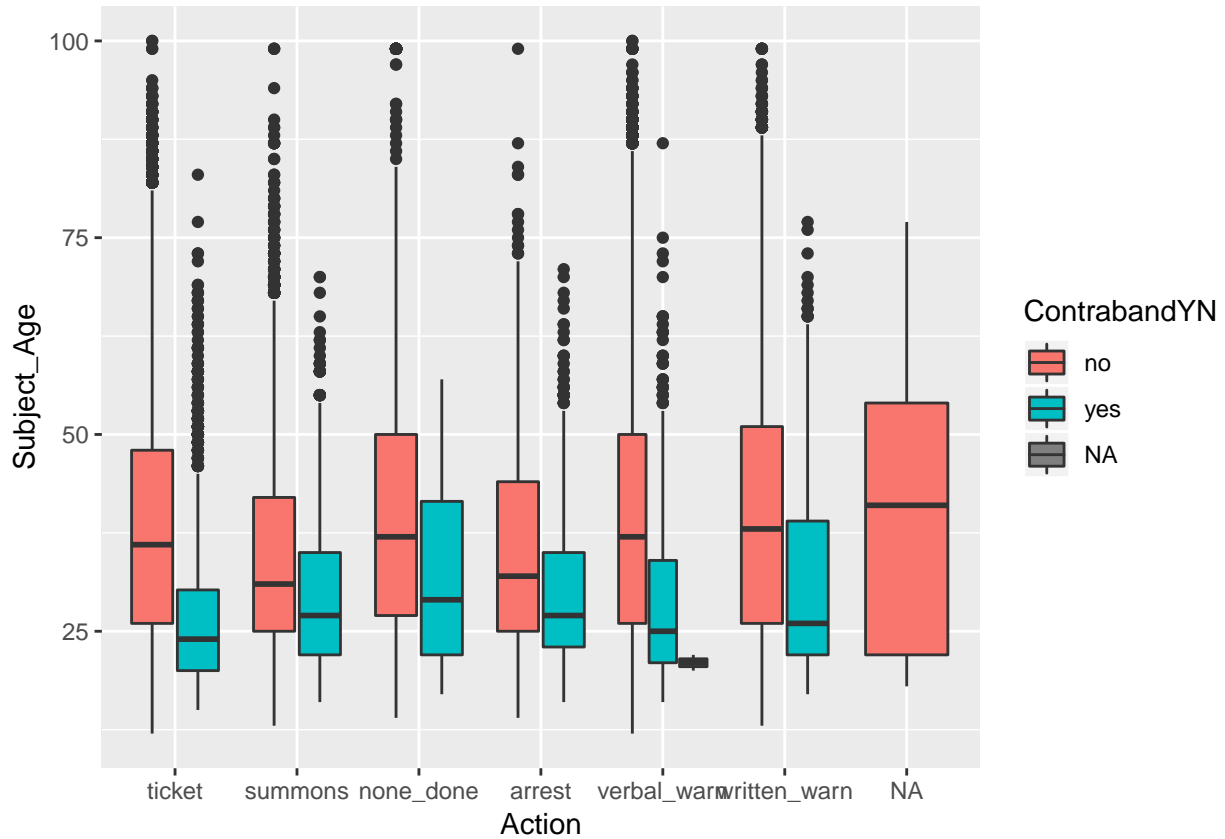
This table and graph are showing what was done by the police officer to conclude the traffic stop. As we can see a ticket is the most common outcome.

Bivariate comparison and Conclusion.

```r
ggplot(Traffic_clean4, aes(x=Dispostion_Cat, y=Subject_Age, fill=ContrabandYN)) +
  geom_boxplot() + xlab("Action")
```

Warning: Removed 1151 rows containing non-finite values (stat_boxplot).

Here we are looking at the 3 variables combined. Our boxplots are grouped by Action then internally split by where or not there was contraband found. All catagories are measured by ages. This shows us that ether younger people are more likely to have contraband, or more likely to get caught caught with it, as our mean of people stopped with contraband is consistantly and significantly lower then those stopped without contraband. Its also interseting to note that among those who were found to have contraband the catagory of ticket has the youngest mean.