Project Part 1

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

## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## v purrr   0.3.4

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

##
## Attaching package: 'gridExtra'

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

##   released colour year age    sex employed citizen checks
## 1      Yes  White 2002  21   Male      Yes     Yes      3
## 2       No  Black 1999  17   Male      Yes     Yes      3
## 3      Yes  White 2000  24   Male      Yes     Yes      3
## 4       No  Black 2000  46   Male      Yes     Yes      1
## 5      Yes  Black 1999  27 Female      Yes     Yes      1
## 6      Yes  Black 1998  16 Female      Yes     Yes      0
```

Data Set Description

Background information

This crime data is a subset of a larger dataset that was collected by the Toronto Star newspaper. It shows data from 1997-2002 on arrests for Marijuana possession in Toronto, Canada. The Toronto star newspaper collected this data to show the differing levels of the city's police department's treatment towards arrestees. The Star used the data to conduct an in-depth investigation of the differences in treatment of blacks and whites by the Toronto police. The data was used by the newspaper to argue that blacks arrested by Toronto police are treated more harshly than whites. (6)

Rows and Variables Explanation

The columns that will be described are only the ones that are relevant to the graphical representation and to the conclusions that can be drawn from it.Checks are the number of times in which the arrestee's name appeared in other police department databases (through previous arrests, previous convictions, parole status, etc.) and shows the criminal history of an arrestee which could help with identifying differing levels of treatment by the Toronto police for blacks and whites. The year column specifies the year when the arrestee was arrested for marijuana possession and can help solidify the trend of levels of treatment for blacks and whites by the police department. The sex and colour columns specifies the sex and the color of the arrestee which is absolutely needed for showcasing differing levels of treatment for the two groups. The citizens column records the citizen status of the arrestee in Canada. The employed column records the employment status of the arrestee and attempts to go into detail about the economic circumstances of the arrestee.

Data Collection Method

The Toronto Star newspaper collected its data from Toronto's police database which stores records of all of the incidents of its arrests, ticketing, etc for all of its arrestees along with some of their background information. This data is a sample of the arrestees for marijuana possession from 1997-2002. The data was sampled by Michael Friendly from York University who randomly selected 1000 arrestees from the data that was gathered from Toronto Star newspaper and therefore a simple random sample was conducted to gather this data. (6)

Potential Issues

A potential issue with this data is that it doesn't exactly go into greater detail for the economic conditions for the arrestees. Employed individuals who are in lower economic classes, because they're not paid much money in their job, may not be able to make enough money to support themselves and their family and therefore may have to turn to crime in order to be able to make money for themselves. One of these ways could even be through dealing drugs which could result in frequent arrests for marijuna possession. Additionally, another issue with this data is that non-citizens of Canada who are arrested will probably not show accurate records of their previous arrests. Toronto's police database can only collect data on individuals who were arrested by Toronto's police and they don't have access to the police databases of other countries. Therefore, the number of checks for non-citizens of could be lower than they really should be, as some non-citizens would've been arrested for crimes in another country, but may have only been arrested once in Canada and only that database would show that individual's name.

Conclusions

Overall, the data shows that black males have the highest average number of checks from 1997-2002. This can be shown by the fact that all of the bars from 1997-2002 for the black males bar graph are higher than than the corresponding bars for the other bar graphs. Some of the factors that may lead to this data could be the economic conditions for black males, or it may be the case that the Toronto police department is more harsh towards black males and therefore this could lead to more arrests. Furthermore, the data also shows that black females had a higher average number of checks than white females from 1997-2002. This can be shown by the fact that all of the bars from 1997-2002 for the black females are higher than the corresponding bars for the white females. Some of the factors that could lead to this data are the economic conditions for black females, or that the police are more harsh towards blacks females than white females which could lead to more arrests for them. (6)

Data Set Graphical and Numerical Summaries

```
arrestData <- data.frame(Arrests)
whiteMaleData <- data.frame(filter(arrestData, colour=="White" & sex =="Male"))
whiteFemaleData <- data.frame(filter(arrestData, colour == "White" & sex
== "Female"))
```

```
blackMaleData <- data.frame(filter(arrestData, colour == "Black" & sex
== "Male"))
blackFemaleData <- data.frame(filter(arrestData, colour == "Black" & sex
== "Female"))

whiteMaleBar <- ggplot(whiteMaleData, aes(x=factor(year), y=checks))+
stat_summary(fun="mean", geom="bar", color="blue", fill = "white" )+
theme_bw()+labs(title="White Males", y = "Average Number of Checks", x = "year")+
theme(axis.title.y = element_text(size=8),  plot.title=
element_text(hjust=0.5, size=11))+expand_limits(y=c(0,2.5)) #(1),(4),(5)

whiteFemaleBar <- ggplot(whiteFemaleData, aes(x=factor(year), y=checks))+
stat_summary(fun="mean", geom="bar", color="pink", fill="white")+theme_bw()+
labs(title="White Females", y = "Average Number of Checks", x= "year")+
theme(axis.title.y = element_text(size=8), plot.title=
element_text(hjust=0.5, size=11))+expand_limits(y=c(0,2.5)) #(1),(4),(5)

blackMaleBar <- ggplot(blackMaleData, aes(x=factor(year), y=checks))+
stat_summary(fun="mean", geom="bar", color="blue", fill="black")+theme_bw()+
labs(title="Black Males", y = "Average Number of Checks", x="year")+
theme(axis.title.y = element_text(size=8), plot.title=
element_text(hjust=0.5, size=11))+expand_limits(y=c(0,2.5)) #(1),(4),(5)

blackFemaleBar <- ggplot(blackFemaleData, aes(x=factor(year), y=checks))+
stat_summary(fun="mean", geom="bar", color="pink", fill="black")+theme_bw()+
labs(title="Black Females", y = "Average Number of Checks", x= "year")+
theme(axis.title.y = element_text(size=8), plot.title=
element_text(hjust=0.5, size=11))+expand_limits(y=c(0,2.5)) #(1),(4),(5)

#(2), (3)
grid.arrange(blackMaleBar, blackFemaleBar,whiteMaleBar, whiteFemaleBar,
top = "Marijuna Possession Arrests in Toronto: Average Number of Checks")
```
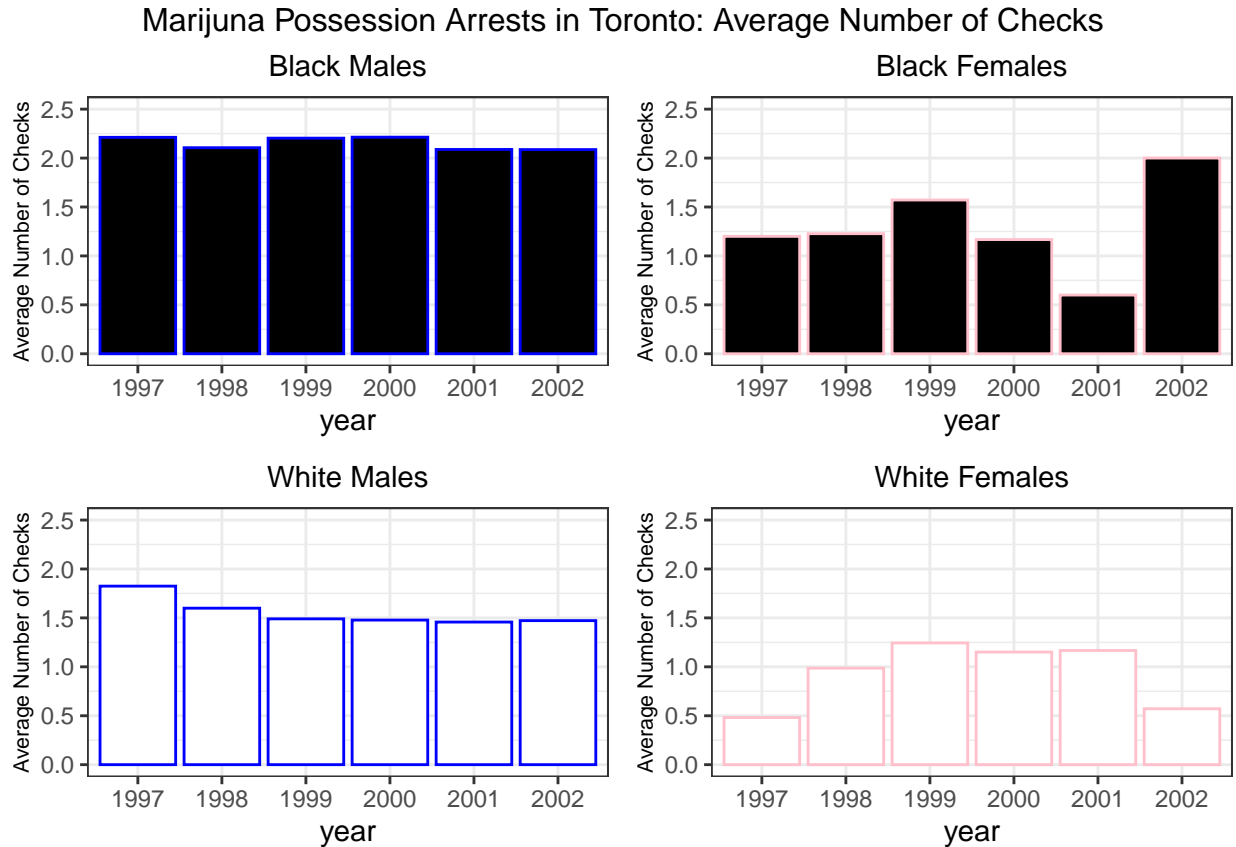
Marijuna Possession Arrests in Toronto: Average Number of Checks

References

1)https://groups.google.com/forum/#!topic/ggplot2/Sz1lMjOI5JA

2)https://cran.r-project.org/web/packages/egg/vignettes/Ecosystem.html

3)https://stackoverflow.com/questions/14726078/changing-title-in-multiplot-ggplot2-using-grid-arrange

4)https://rstudio-pubs-static.s3.amazonaws.com/366989_b80cfe1ffa5149f589fb44ea847f5967.html

5)http://www.sthda.com/english/wiki/ggplot2-axis-scales-and-transformations

6)https://www.thestar.com/news/gta/knowntopolice/2002/10/19/singled-out.html