Variable Investigation
for
Kelowna Weather-Crash Project

Contents

1 Loading data

```
> library(ggplot2)
> library(ggthemes)
> theme_set(theme_few())
> library(tidyverse)
> load_first_object <- function(fname){
+   #this function was written by Dr. Rhonda Rosychuk at the U of A
+   e <- new.env(parent = parent.frame())
+   load(fname, e)
+   return(e[[ls(e)[1]]])
+ }
> weatherdata = load_first_object("../../rda_files/all_data.rda")
```

2 Accidents over Time

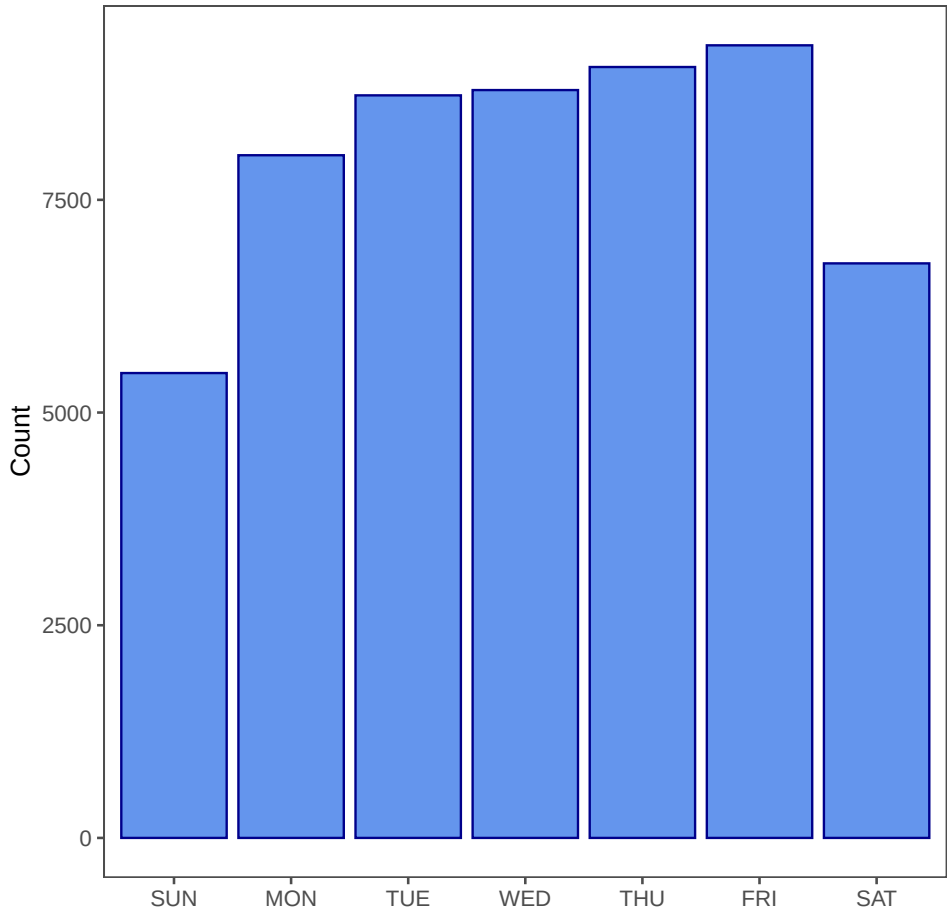
2.1 Day of the Week

Distribution of accidents throughout the week:

```
> #reordering factor
> weeknames = c("SUNDAY", "MONDAY", "TUESDAY", "WEDNESDAY", "THURSDAY", "FRIDAY", "SATURDAY"
> weatherdata$Day.Of.Week = factor(weatherdata$Day.Of.Week,
+                                  levels=weeknames)
> weatherdata$Month.Of.Year = factor(weatherdata$Month.Of.Year,
+                                    levels=toupper(month.name))
> table(weatherdata$Day.Of.Week)

   SUNDAY    MONDAY   TUESDAY WEDNESDAY  THURSDAY    FRIDAY  SATURDAY
     5464      8024      8728      8790      9061      9316      6753

> weatherdata %>%
+   ggplot(aes(x=Day.Of.Week)) +
+   geom_histogram(stat='count', colour='#00008b', fill='#6495ed') +
+   xlab('') +
+   ylab('Count') +
+   scale_x_discrete(labels=c(substr(weeknames, start=1, stop=3)))
```

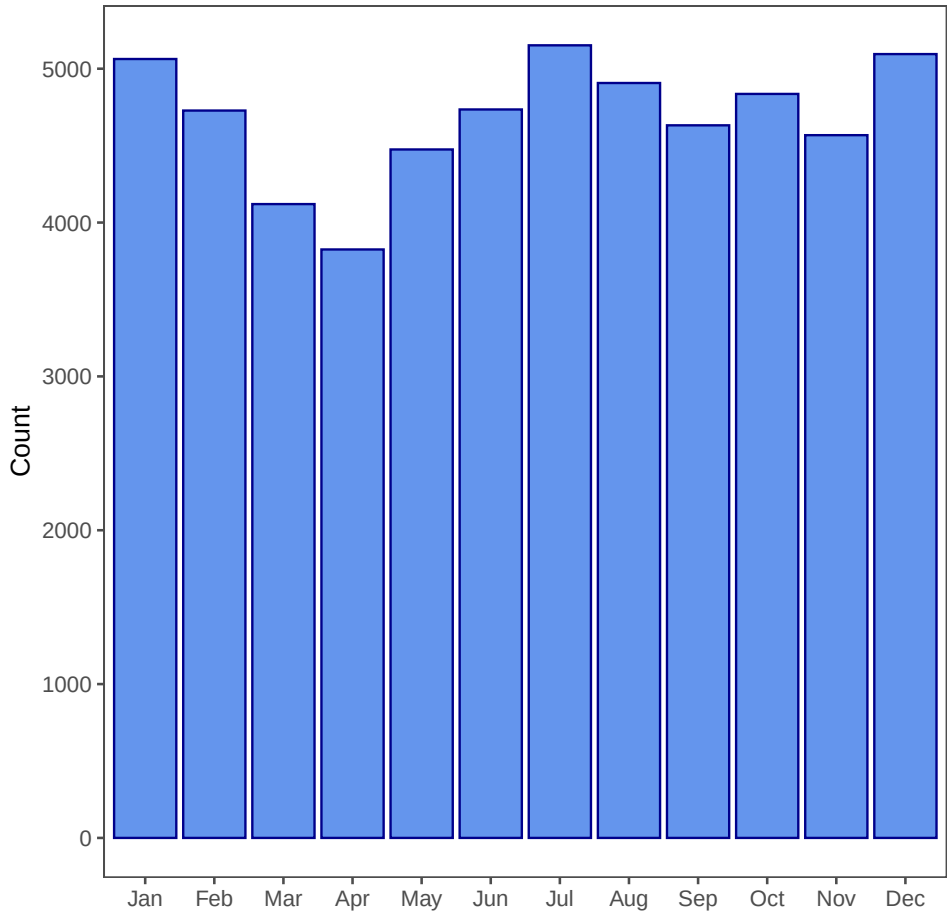


2.2 Month of the Year

```
> #making monthnumber column
> weatherdata[,"monthnumber"] = match(tolower(weatherdata$Month.Of.Year),
+                                     tolower(month.name))
> table(weatherdata$Month.Of.Year)

  JANUARY  FEBRUARY     MARCH     APRIL       MAY      JUNE      JULY    AUGUST
     5063      4728      4120      3825      4475      4735      5152      4907
SEPTEMBER   OCTOBER  NOVEMBER  DECEMBER
     4632      4836      4568      5095

> weatherdata %>%
+   ggplot(aes(x=Month.Of.Year)) +
+   geom_histogram(stat='count', colour='#00008b', fill='#6495ed') +
+   xlab('') +
+   ylab('Count') +
+   scale_x_discrete(labels=(month.abb))
```

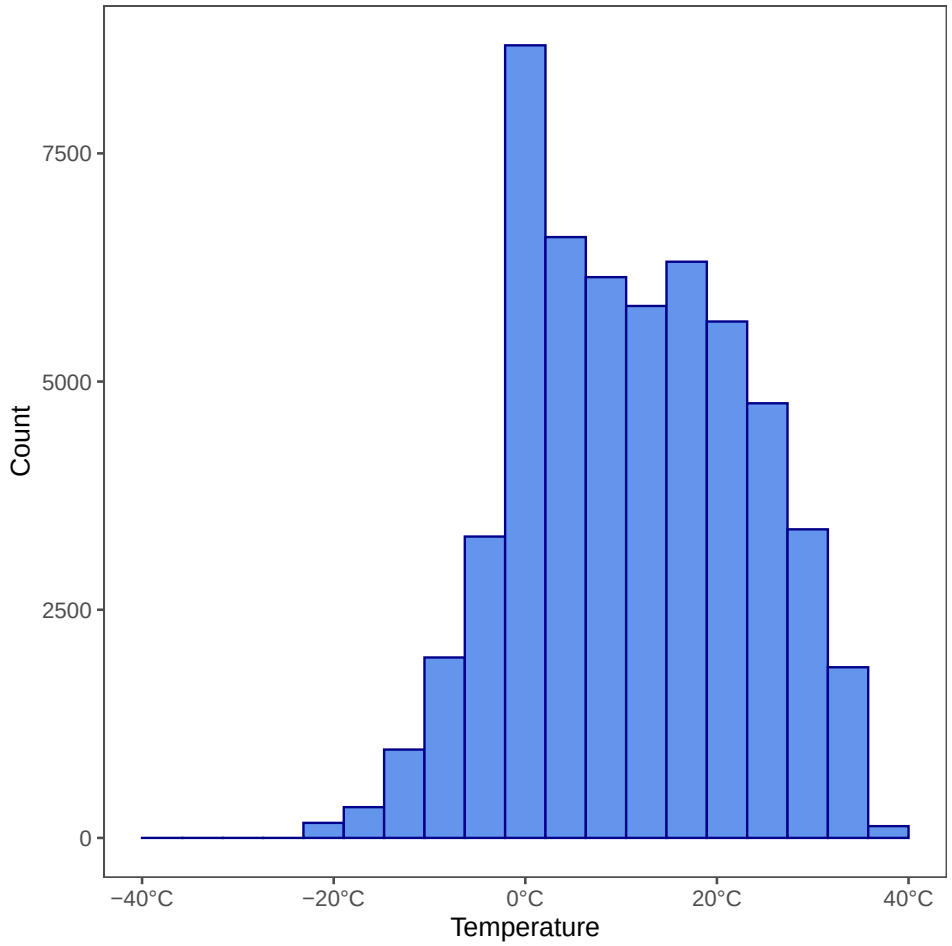


3 Temperature

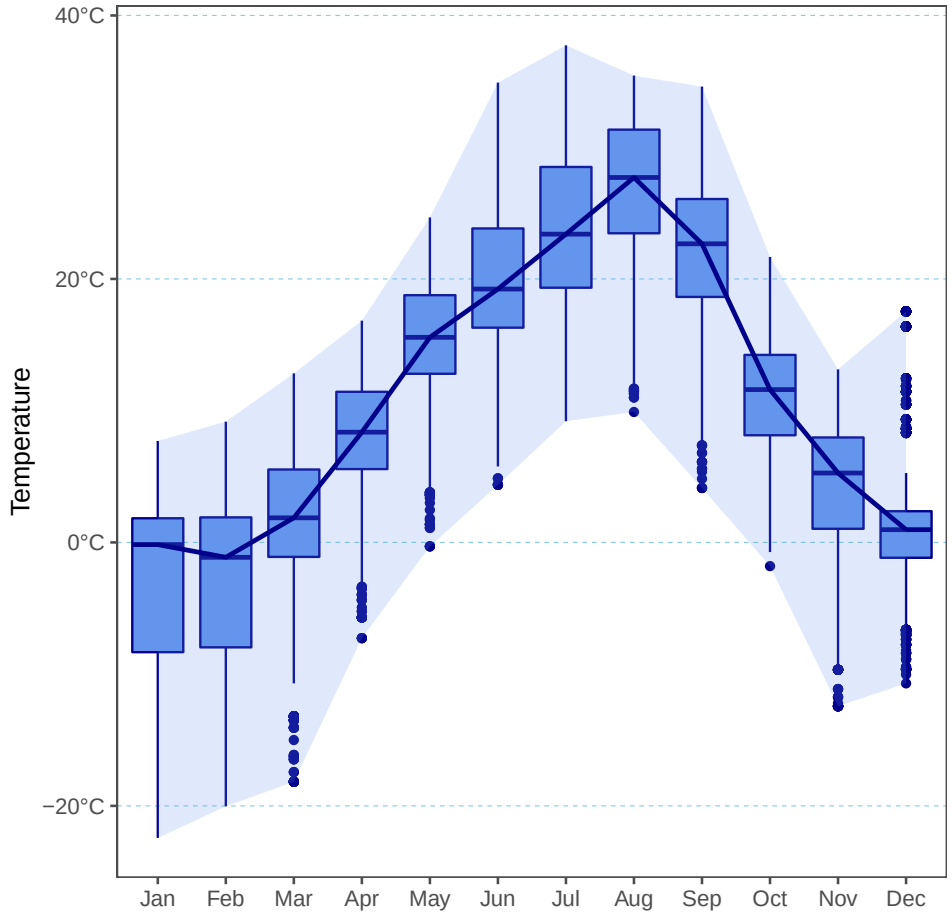
```
> summary(weatherdata$Temp...C.)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 -22.43    1.60   10.37   10.92   20.00   37.73      35

> weatherdata %>%
+   ggplot(aes(x=Temp...C.)) +
+   geom_histogram(colour='#00008b', fill='#6495ed', bins=20) +
+   xlab('Temperature') +
+   ylab('Count') +
+   scale_x_continuous(labels = scales::label_number(suffix = "°C", accuracy=1),
+                      limits = c(-40, 40))
```




```
> weatherdata %>%
+   ggplot(aes(x=monthnumber, y=Temp...C.)) +
+   geom_boxplot(aes(x=Month.Of.Year), colour='#00008b', fill='#6495ed') +
+   geom_smooth(stat = 'summary', alpha = 0.2, fill = '#6495ed', color = '#00008b',
+                fun.data = median_hilow, fun.args = list(conf.int = 1)) +
+   #geom_line(stat='summary', fun='median', color='#00008b', lwd=1.2) +
+   xlab('') +
+   ylab('Temperature') +
+   scale_x_discrete(labels=month.abb) +
+   scale_y_continuous(labels = scales::label_number(suffix = "°C")) +
+   theme(panel.grid.major.y = element_line(color = "#8ccde3",
+                                   size = 0.25,
+                                   linetype = 2))
```


40°C
20°C
0°C
−20°C
Temperature
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

4 Crash Severity