

---

VARIABLE INVESTIGATION  
FOR  
KELOWNA WEATHER-CRASH PROJECT

## Contents

<b>1</b>	<b>Loading data</b>	<b>2</b>
<b>2</b>	<b>Accidents over Time</b>	<b>2</b>
2.1	Day of the Week . . . . .	2
2.2	Month of the Year . . . . .	3
<b>3</b>	<b>Temperature</b>	<b>4</b>
<b>4</b>	<b>Crash Severity</b>	<b>7</b>

# 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]]])
+ }
> #cleaned and combined
> alldata = load_first_object("../rda_files/all_data.rda")
> #weatherdata
> weatherdata = c()
> for (i in c(2017:2021)){
+   for (j in c(1:12)){
+     temp = subset(read.csv(paste0("../weatherdata/en_climate_hourly_BC_1123939_",
+       sprintf("%02d", j), '-', i, '_P1H.csv')),
+       select = - c(`Temp.Flag`,
+         `Dew.Point.Temp.Flag`, `Rel.Hum.Flag`,
+         `Precip..Amount.Flag`, `Wind.Dir.Flag`,
+         `Wind.Spd.Flag`, `Visibility.Flag`,
+         `Stn.Press.Flag`, `Hmdx`, `Hmdx.Flag`, `Wind.Chill.Flag`))
+     weatherdata = rbind(weatherdata, temp)
+   }
+ }
```

# 2 Accidents over Time

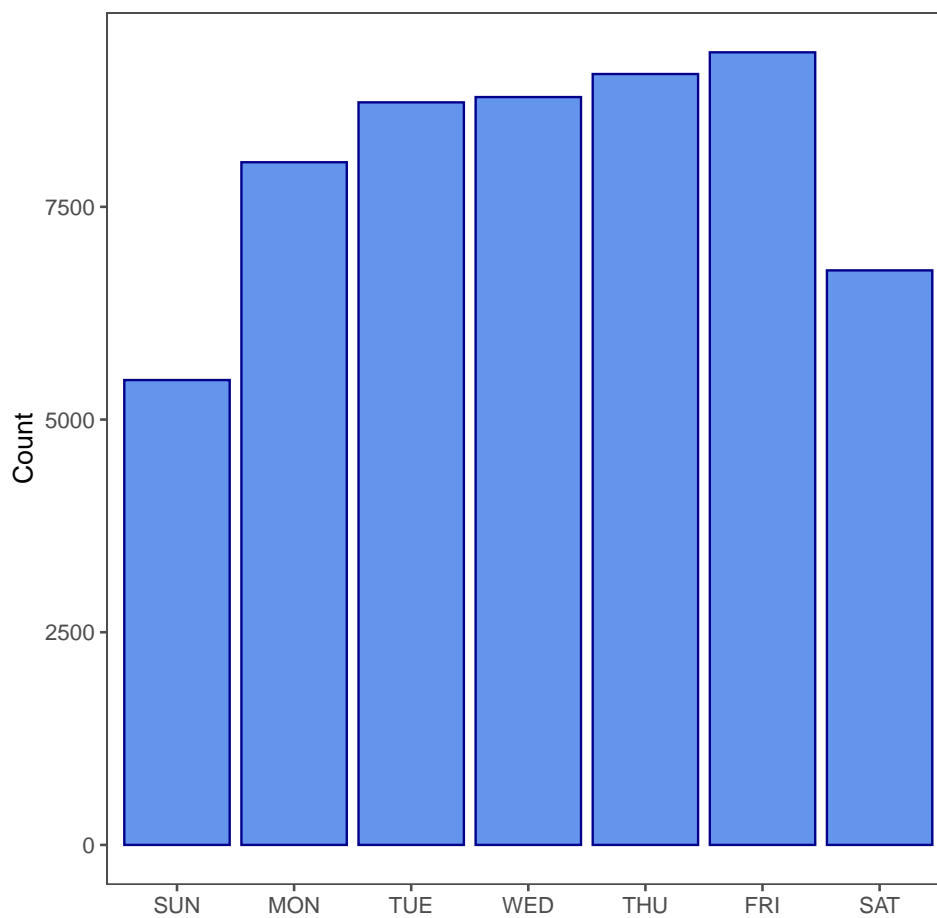
## 2.1 Day of the Week

Distribution of accidents throughout the week:

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

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

```
> alldata %>%
+   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)))
```

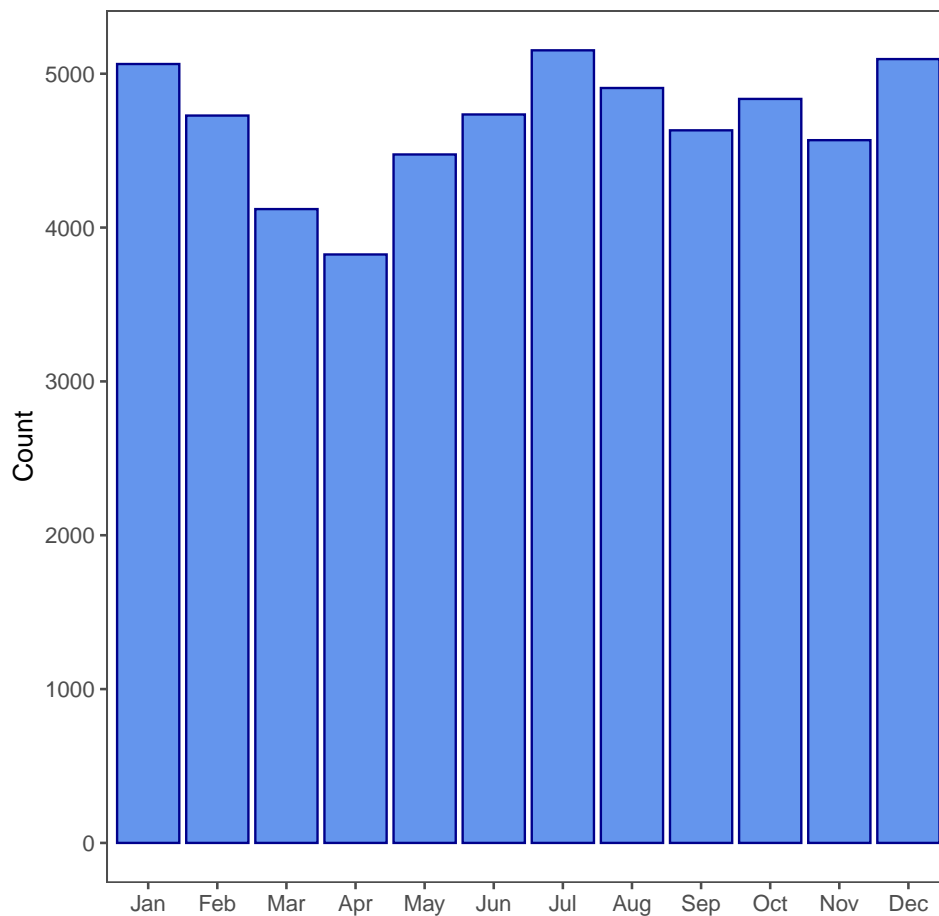


## 2.2 Month of the Year

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

```
> alldata %>%
+   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

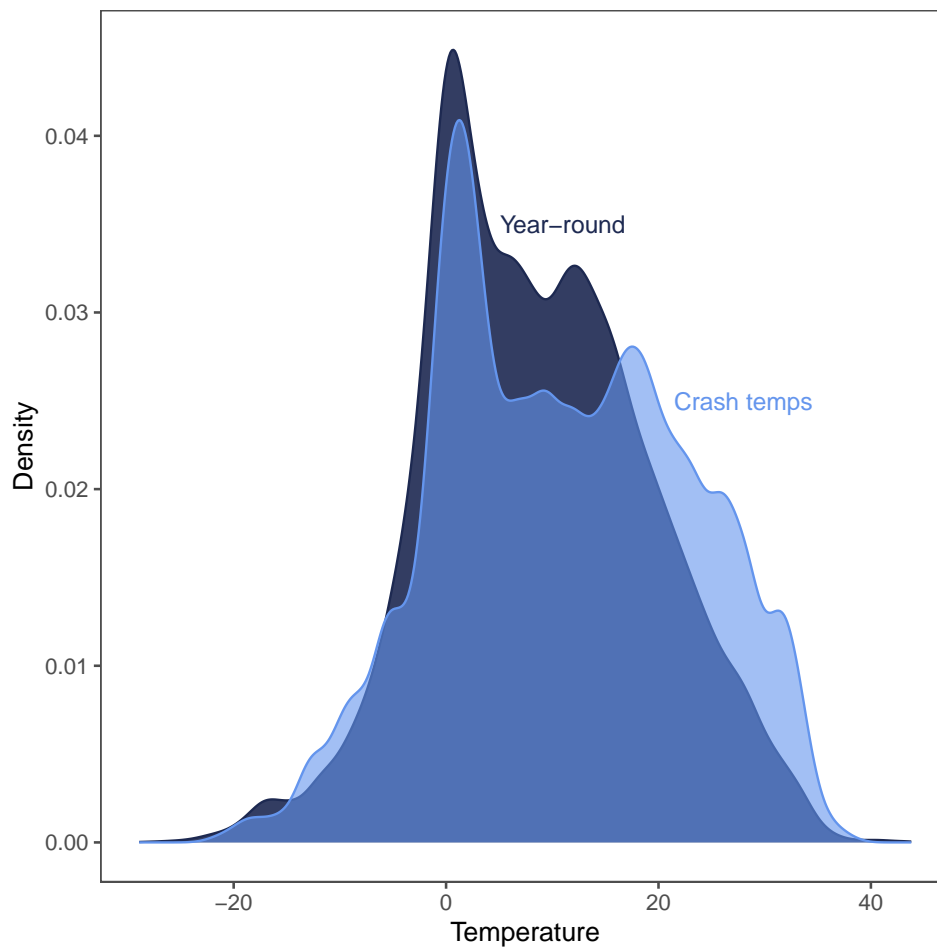
```
> summary(alldata$Temp...C.)
> alldata %>%
+   ggplot(aes(x=Temp...C.)) +
+   geom_histogram(colour='#00008b', fill='#6495ed', bins=20) +
+   xlab('Temperature') +
+   ylab('Count') +
+   scale_x_continuous(labels = scales::label_number(suffix = "°C", accuracy=1),
+                       limits =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

> data.frame(
+   'Avg temp.' = mean(weatherdata$Temp...C., na.rm=TRUE),
+   'Avg crash temp.' = mean(alldata$Temp...C., na.rm = TRUE)
+ )

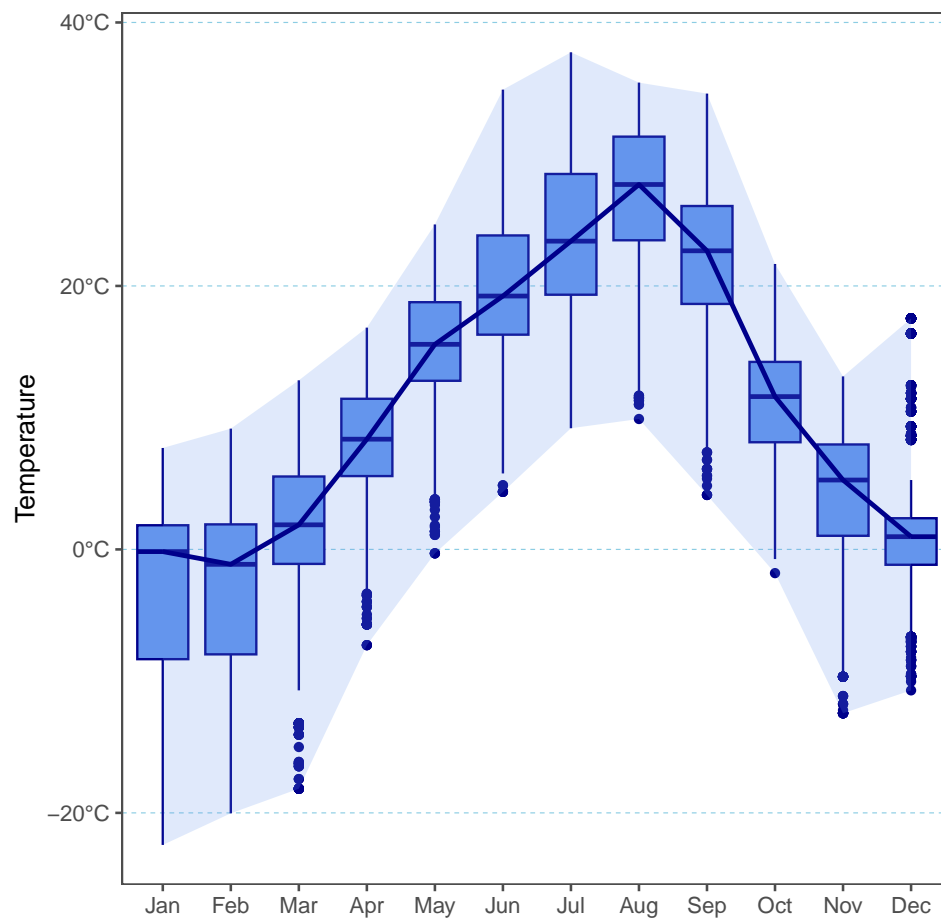
  Avg.temp. Avg.crash.temp.
1  8.511442    10.91872

> ggplot(data=weatherdata, aes(x=Temp...C.)) +
+   geom_density(colour='#1d2951', fill='#1d2951', alpha=0.9) +
+   geom_density(data=alldata, colour='#6495ed', fill='#6495ed', alpha=0.6) +
+   xlab('Temperature') +
+   ylab('Density') +
```

```
+ annotate(
+   geom="text", x=c(11, 28), y=c(0.035, 0.025), label=c("Year-round", "Crash temps"), col
+ )
```



```
> alldata %>%
+   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))
```



## 4 Crash Severity