

Tweet Count Data Analysis

Kyle Krakoski

12/15/2020

```
library(ggplot2)
## Warning: package 'ggplot2' was built under R version 4.0.3
library(dplyr)
## Warning: package 'dplyr' was built under R version 4.0.3
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library(ggformula)
## Warning: package 'ggformula' was built under R version 4.0.3
## Loading required package: ggstance
## Warning: package 'ggstance' was built under R version 4.0.3
##
## Attaching package: 'ggstance'
## The following objects are masked from 'package:ggplot2':
##
##   geom_errorbarh, GeomErrorbarh
##
## New to ggformula? Try the tutorials:
##   learnr::run_tutorial("introduction", package = "ggformula")
##   learnr::run_tutorial("refining", package = "ggformula")
library(tidyverse)
## Warning: package 'tidyverse' was built under R version 4.0.3
```

```
## -- Attaching packages -----
----- tidyverse 1.3.0 --

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

## Warning: package 'tidyr' was built under R version 4.0.3
## Warning: package 'purrr' was built under R version 4.0.3
## Warning: package 'forcats' was built under R version 4.0.3

## -- Conflicts -----
----- tidyverse_conflicts() --
## x dplyr::filter()          masks stats::filter()
## x ggstance::geom_errorbarh() masks ggplot2::geom_errorbarh()
## x dplyr::lag()             masks stats::lag()

library(ggpubr)

## Warning: package 'ggpubr' was built under R version 4.0.3

library(rstatix)

## Warning: package 'rstatix' was built under R version 4.0.3

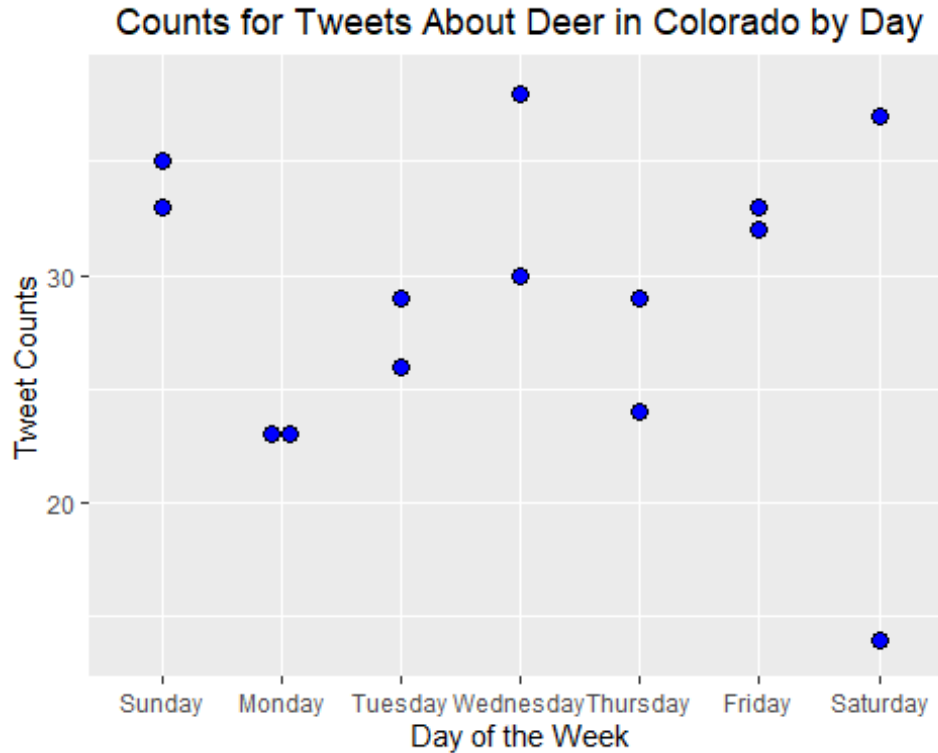
##
## Attaching package: 'rstatix'

## The following object is masked from 'package:stats':
##
##      filter

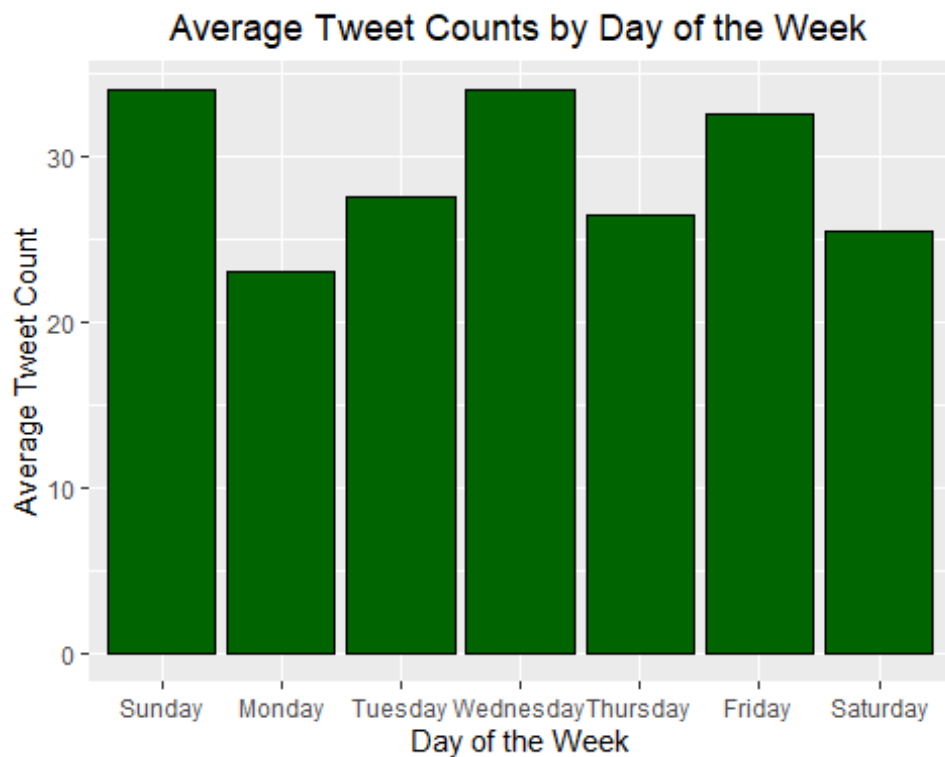
daily_tweet_counts = read.csv('Tweet_Count_By_Day_for_R.csv',
fileEncoding="UTF-8-BOM")
tweet_counts_df = data.frame(daily_tweet_counts, na.omit = TRUE) #Creates a
dataframe, ignores NA values

reorder_Days <- factor(tweet_counts_df$Day, level = c('Sunday', 'Monday',
'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday')) #Used to reorder
the x-axis, so days display chronologically instead of alphabetically
ggplot(data = tweet_counts_df, aes(x=reorder_Days, y= Count))+ #Creates a
scatterplot of the raw data
  labs(title = 'Counts for Tweets About Deer in Colorado by Day', x = 'Day of
the Week', y = 'Tweet Counts')+
  theme(plot.title = element_text(hjust = 0.5))+
  geom_dotplot(binaxis = "y", stackdir = "center", color = "black", fill =
"blue") #Since Monday had the same count for both samplings, reformatted the
scatterplot to show equal tweet counts side by side.
```

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



```
avg_count_df = data.frame(aggregate(Count~Day,data=tweet_counts_df,FUN=mean))  
#Computes the average tweet count by day of the week and sets it to a  
#dataframe  
avg_count_df$Day = reorder_Days <- factor(avg_count_df$Day, level =  
c('Sunday', 'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday',  
'Saturday')) #Reorders day of the week chronologically  
  
gf_col(Count~reorder_Days, data = avg_count_df, color = "black", fill =  
"darkgreen")+ #Creates a bar graph of the average count data  
  labs(title = 'Average Tweet Counts by Day of the Week', x = 'Day of the  
Week', y = 'Average Tweet Count')+  
  theme(plot.title = element_text(hjust = 0.5))
```



```
anova_test <- tweet_counts_df %>% anova_test(Count ~ Day) #ANOVA test on my
data

## Coefficient covariances computed by hccm()

anova_test

## ANOVA Table (type II tests)
##
##   Effect DFn DFd      F      p p<.05 ges
## 1    Day    6    7 0.879 0.555      0.43

tukey_test <- tweet_counts_df %>% tukey_hsd(Count ~ Day) #Unnecessary since
ANOVA test failed to reject the null hypothesis, but included as it would be
the next step if null hypothesis was rejected.
tukey_test

## # A tibble: 21 x 9
##   term group1 group2 null.value estimate conf.low conf.high p.adj
p.adj.signif
## * <chr> <chr> <chr>      <dbl>      <dbl>      <dbl>      <dbl> <dbl> <chr>
## 1 Day   Friday Monday      0      -9.5      -36.1      17.1 0.781 ns
## 2 Day   Friday Satur~      0      -7.      -33.6      19.6 0.927 ns
## 3 Day   Friday Sunday      0       1.5     -25.1      28.1 1      ns
## 4 Day   Friday Thurs~      0      -6.      -32.6      20.6 0.963 ns
## 5 Day   Friday Tuesd~      0      -5.      -31.6      21.6 0.984 ns
## 6 Day   Friday Wedne~      0      1.50     -25.1      28.1 1      ns
## 7 Day   Monday Satur~      0       2.5     -24.1      29.1 1      ns
```

| | | | | | | | | |
|----|--------|-----------------------|---|------|-------|------|-------|----|
| ## | 8 Day | Monday Sunday | 0 | 11. | -15.6 | 37.6 | 0.669 | ns |
| ## | 9 Day | Monday Thurs~ | 0 | 3.50 | -23.1 | 30.1 | 0.997 | ns |
| ## | 10 Day | Monday Tuesd~ | 0 | 4.5 | -22.1 | 31.1 | 0.991 | ns |
| ## | # | ... with 11 more rows | | | | | | |