ST565: Time Series HW1

Amirhosein "Emerson" Azarbakht azarbaka@oregonstate.edu

Q1

Google maintains records of the popularity of search terms over time at http://www.google.com/trends You can examine daily data by selecting a time period of less than 90 days, otherwise you'll see weekly or monthly data (e.g. searches for "hangover" in the last 90 days). The numbers reported are relative search volume. Your task is to find:

- one search term dominated by a weekly or annual seasonal pattern:
- Oktoberfest
- one search term dominated by an increasing trend:
- US Presidential Elections
- one search term that would be poorly described by a combination of trend and seasonality:
- Climate Change

In each case, you should download the relevant data as a .csv file (click on the ... on the trends page), read the data into R and produce a plot of the data.

Q1.1. Weekly or Annual Seasonal pattern: Oktoberfest

```
library(ggplot2)
library(dplyr)
```

```
##
## 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(lubridate)
options(stringsAsFactors = FALSE)

setwd("/home/aha/Desktop/Time_Series_ST565/HW1/")
data <- read.csv("Oktoberfest.csv")
str(data)
```

```
## 'data.frame':    627 obs. of  2 variables:
##  $ Week       : chr  "2004-01-04 - 2004-01-10" "2004-01-11 - 2004-01-17" "2004-01-18 - 2004-01-24" "
##  $ oktoberfest: int  3 3 4 3 4 3 3 4 4 5 ...
```

```r
head(data)
```

```
##                      Week oktoberfest
## 1 2004-01-04 - 2004-01-10           3
## 2 2004-01-11 - 2004-01-17           3
## 3 2004-01-18 - 2004-01-24           4
## 4 2004-01-25 - 2004-01-31           3
## 5 2004-02-01 - 2004-02-07           4
## 6 2004-02-08 - 2004-02-14           3
```

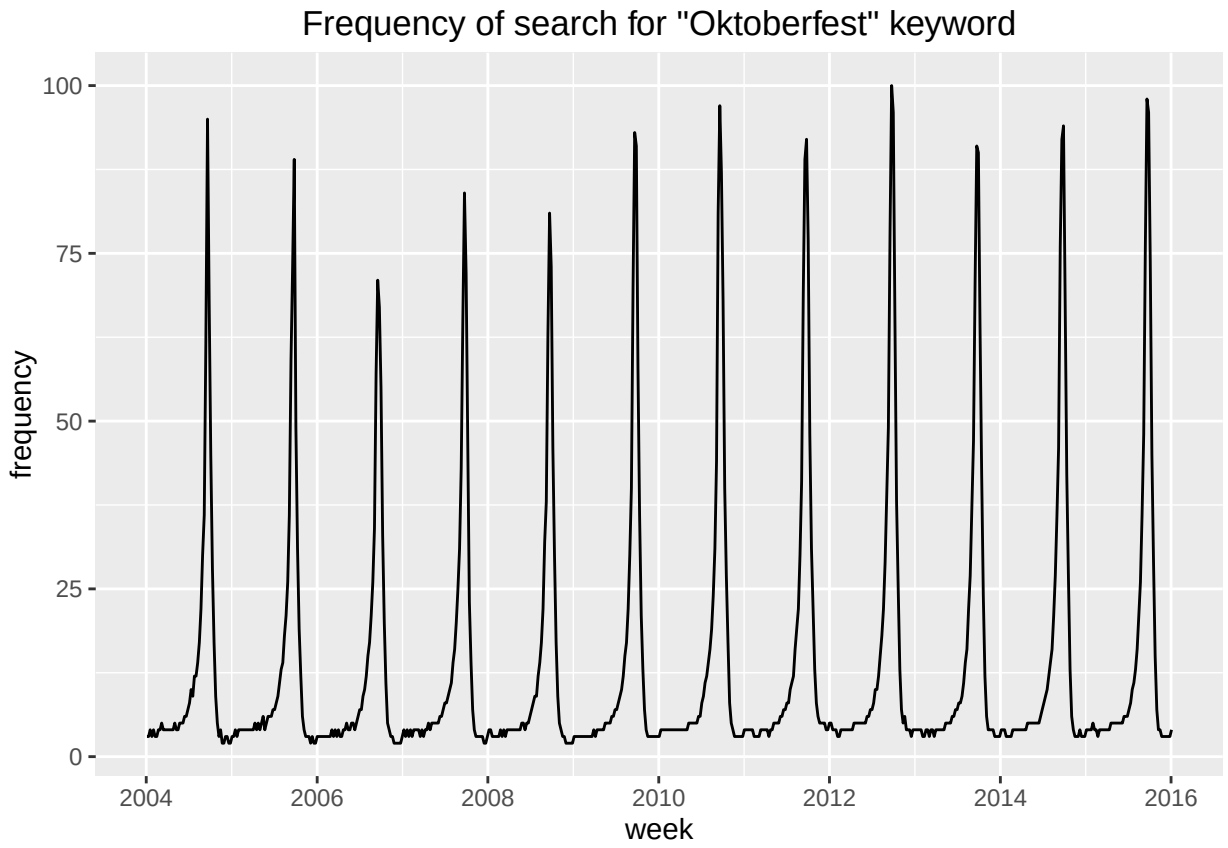
```r
# Parse the dates
#?parse_date_time
#strptime(week, "%Y-%m-%d - %Y-%m-%d")
data$StartDatePOSIX <- strptime(data$Week, "%F")
data$EndDatePOSIX <- strptime(data$Week, "%F - %F")
# parsed the dates as POSIX format

# convert to Date format
data$StartDate <- as.Date(data$StartDatePOSIX)
data$EndDate <- as.Date(data$EndDatePOSIX)

#data$searchFrequency <- data$oktoberfest
data <- data %>% select(StartDate, oktoberfest) %>% rename(week = StartDate, frequency = oktoberfest)
head(data)
```

```
##         week frequency
## 1 2004-01-04         3
## 2 2004-01-11         3
## 3 2004-01-18         4
## 4 2004-01-25         3
## 5 2004-02-01         4
## 6 2004-02-08         3
```

```r
# plot frequency of keyword "Oktoberfest" search
qplot(week, frequency, data = data, geom = "line", main = "Frequency of search for \"Oktoberfest\" keyw
```



Q1.2. An increasing trend: U.S. Presidential Elections

```
data <- read.csv("USPresidentialElections.csv")
str(data)
```

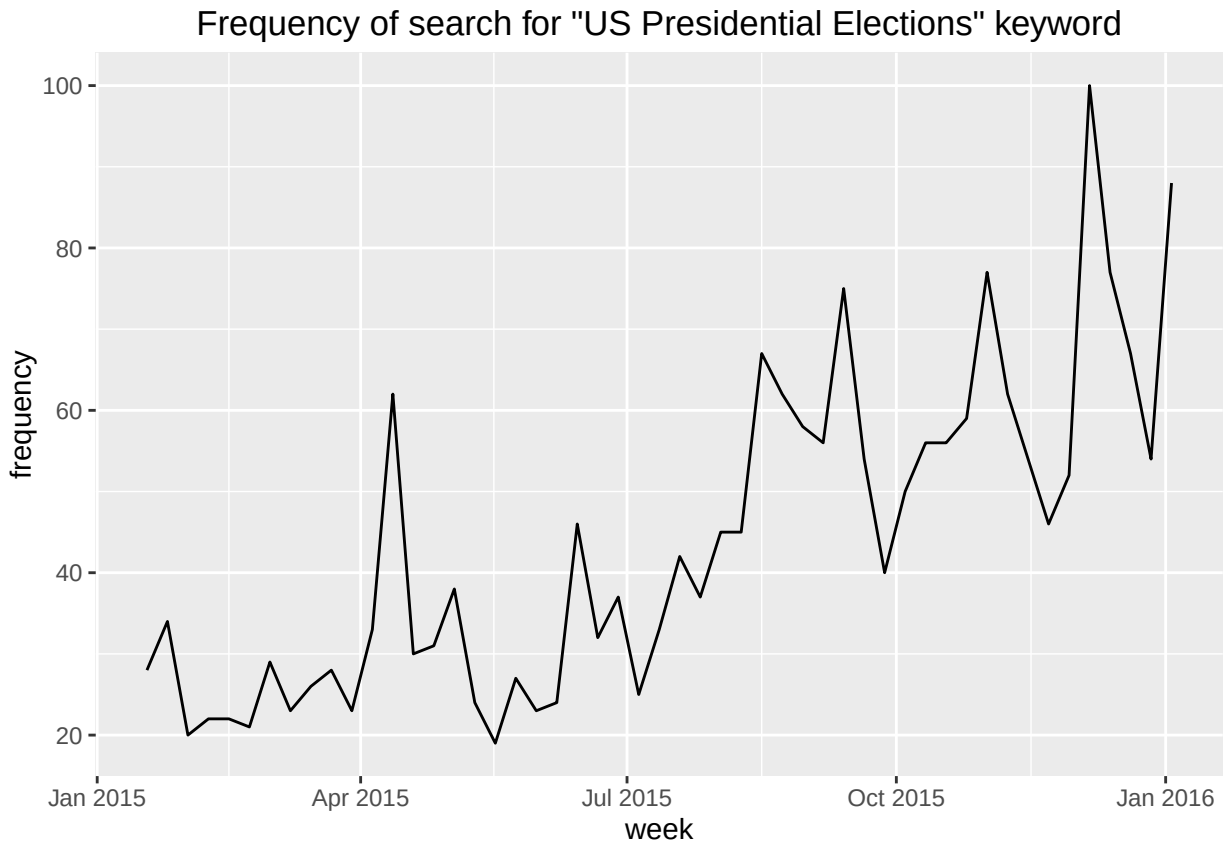
```
## 'data.frame':    51 obs. of  2 variables:
##  $ Week                     : chr  "2015-01-18 - 2015-01-24" "2015-01-25 - 2015-01-31" "2015-02-01 -
##  $ us.presidential.elections: int  28 34 20 22 22 21 29 23 26 28 ...
```

```
head(data)
```

```
##                       Week us.presidential.elections
## 1 2015-01-18 - 2015-01-24                         28
## 2 2015-01-25 - 2015-01-31                         34
## 3 2015-02-01 - 2015-02-07                         20
## 4 2015-02-08 - 2015-02-14                         22
## 5 2015-02-15 - 2015-02-21                         22
## 6 2015-02-22 - 2015-02-28                         21
```

```
data$StartDate <-as.Date(strptime(data$Week, "%F"))
data <- data %>% select(StartDate, us.presidential.elections) %>% rename(week = StartDate, frequency = u

qplot(week, frequency, data = data, geom = "line", main = "Frequency of search for \"US Presidential Ele
```



Q1.3. One search term that would be poorly described by a combination of trend and seasonality: Climate Change

```
data <- read.csv("ClimateChange.csv")
str(data)
```

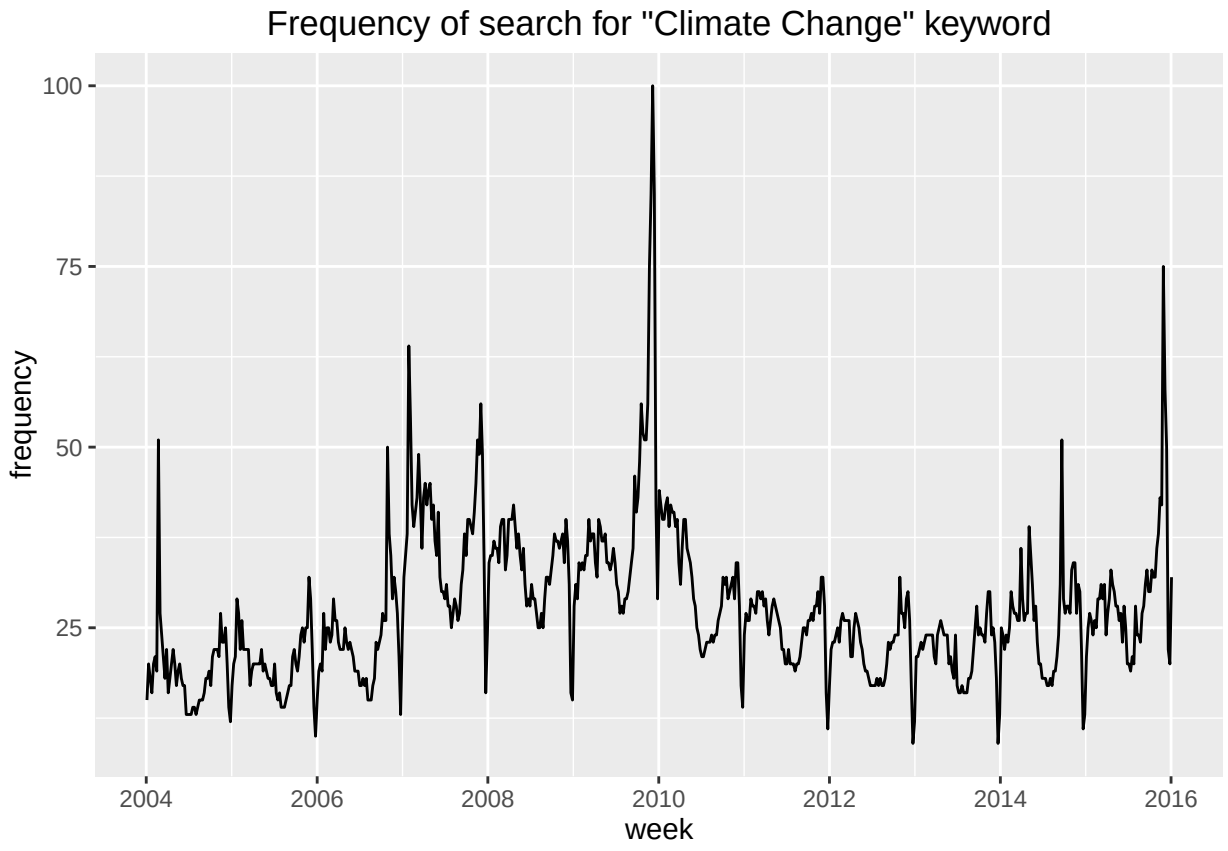
```
## 'data.frame':    627 obs. of  2 variables:
##  $ Week          : chr  "2004-01-04 - 2004-01-10" "2004-01-11 - 2004-01-17" "2004-01-18 - 2004-01-24
##  $ climate.change: int  15 20 18 16 20 21 19 51 27 24 ...
```

```
head(data)
```

```
##                      Week climate.change
## 1 2004-01-04 - 2004-01-10             15
## 2 2004-01-11 - 2004-01-17             20
## 3 2004-01-18 - 2004-01-24             18
## 4 2004-01-25 - 2004-01-31             16
## 5 2004-02-01 - 2004-02-07             20
## 6 2004-02-08 - 2004-02-14             21
```

```
data$StartDate <- as.Date(strptime(data$Week, "%F"))
data <- data %>% select(StartDate, climate.change) %>% rename(week = StartDate, frequency = climate.cha
```

```
qplot(week, frequency, data = data, geom = "line", main = "Frequency of search for \"Climate Change\" k
```



Q2

I've downloaded much higher resolution data from the Corvallis Municipal Airport weather station. I am giving you ten years of about thrice hourly (I think, you should probably check) observations. Download the data from: http://stat565.cwick.co.nz/data/corv_sub.rds.

Check out the structure, I've been nice and already converted the date and datetime to R date time classes.

Before I move to a new place I like to look up what the climate is like there. But I normally end up looking at a graph that shows me the average min, mean and max temperature by month, and maybe total precipitation, and that doesn't really help me decide how my day to day life will be affected by the weather. I'm much more interested in things like:

- How often will I be biking to work in below freezing temperatures?
- How many days a year will I see the sun?
- How many days do I need a raincoat, assuming I'm only outside during my morning and evening commute?

Your task is to create a climate metric that you would find useful, and by using the data provided and dplyr, calculate it and present it visually. (And since this is a statistics class you might give a thought to how you would quantify and present the variability in your metric).

Q2.1. When do heaviest percipitation happen in a day? (e.g. early morning? noon? evening?)

```
corv <- readRDS("corv_sub.rds")
head(corv)
```

```
##              datetime       date hour temp precip    conditions
## 1 2005-01-01 00:15:00 2005-01-01    0 42.8     NA      Overcast
## 2 2005-01-01 00:35:00 2005-01-01    0 42.8     NA Mostly Cloudy
## 3 2005-01-01 00:55:00 2005-01-01    0 42.8     NA      Overcast
## 4 2005-01-01 01:15:00 2005-01-01    1 42.8     NA Mostly Cloudy
## 5 2005-01-01 01:35:00 2005-01-01    1 42.8     NA Mostly Cloudy
## 6 2005-01-01 01:55:00 2005-01-01    1 42.8     NA      Overcast
```

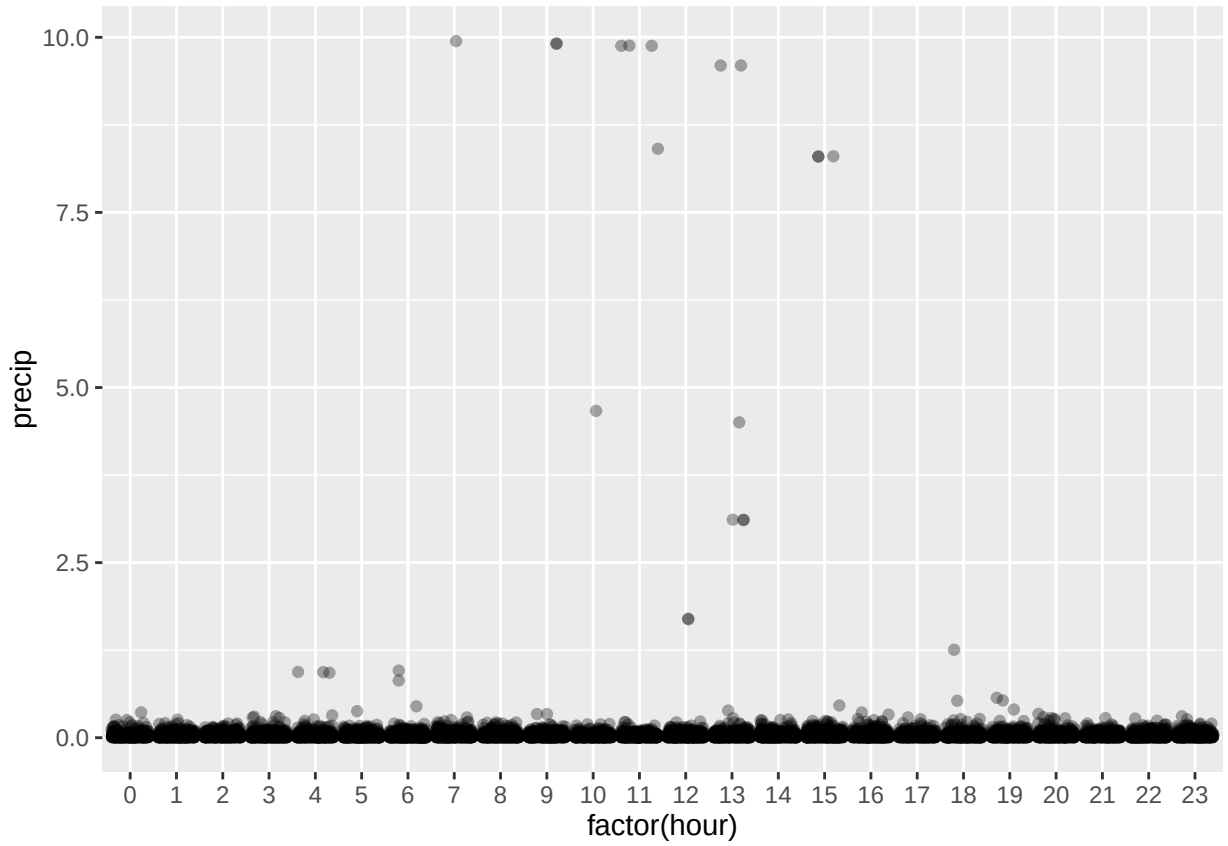
```
str(corv)
```

```
## 'data.frame':    234175 obs. of  6 variables:
##  $ datetime  : POSIXct, format: "2005-01-01 00:15:00" "2005-01-01 00:35:00" ...
##  $ date      : POSIXct, format: "2005-01-01" "2005-01-01" ...
##  $ hour      : int  0 0 0 1 1 1 2 2 2 3 ...
##  $ temp      : num  42.8 42.8 42.8 42.8 42.8 42.8 42.8 42.8 42.8 42.8 ...
##  $ precip    : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ conditions: Factor w/ 7 levels "Thunderstorm",..: 2 3 2 3 3 2 3 3 2 3 ...
```

```
corv <- corv %>% rename(datePOSIX = date)
corv$date <- as.Date(strptime(corv$datePOSIX, "%F"))

# corvallisClean <- filter(coru, !is.na(precip))
# head(corvallisClean)
# attach(corvallisClean)

qplot(factor(hour), precip, data = filter(corv, !is.na(precip)), geom = "jitter", alpha = I(1/3))
```




```
# qplot(factor(hour), temp, data = filter(corv, !is.na(temp)), alpha = I(1/10))
```

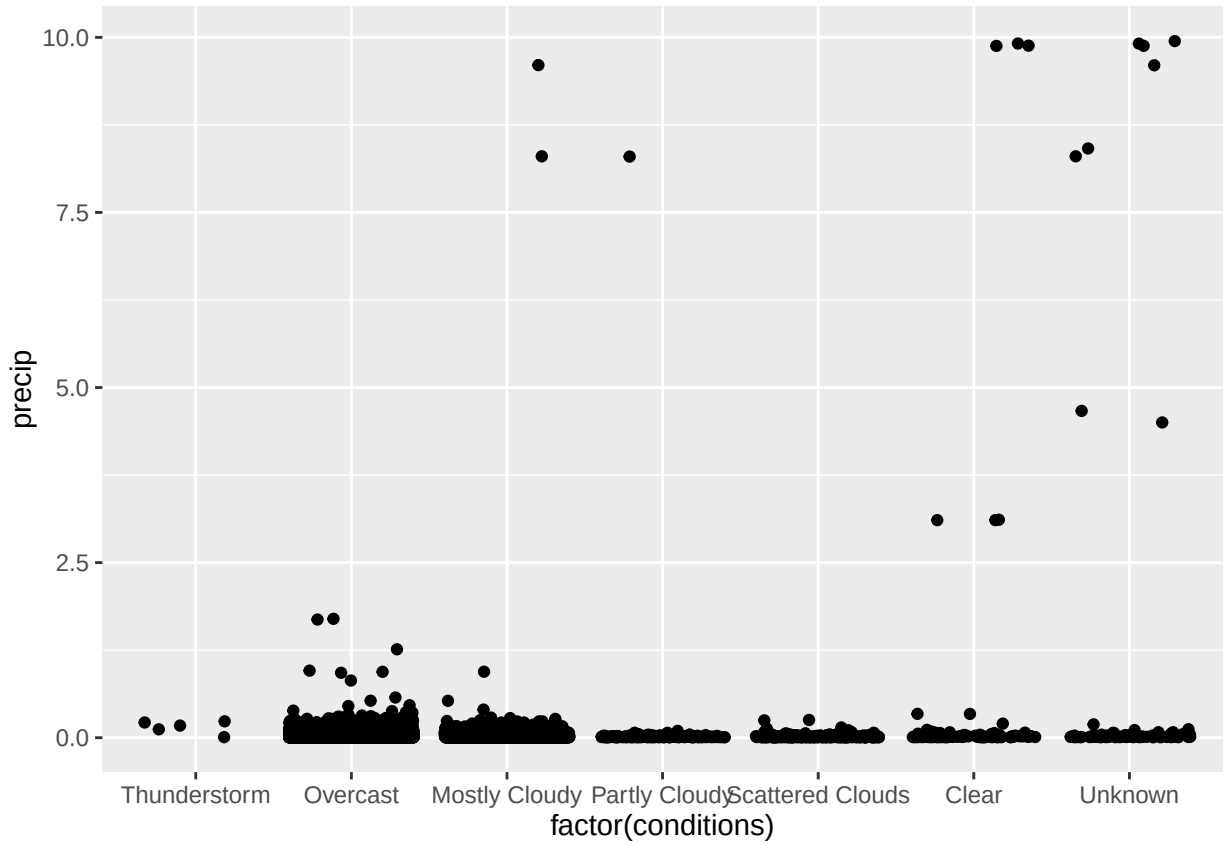
Q2.2. How much precipitation should I expect in a weather condition?

```
attach(corv)
```

```
## The following object is masked from package:datasets:
##
##     precip
```

```
# factor(conditions)

qplot(factor(conditions), precip, data = filter(corv, !is.na(precip)), geom = "jitter")
```



Q2.3. Which days of the year are rainy?

```
attach(corv)
```

```
## The following objects are masked from corv (pos = 3):
##
##     conditions, date, datePOSIX, datetime, hour, precip, temp

## The following object is masked from package:datasets:
##
##     precip
```

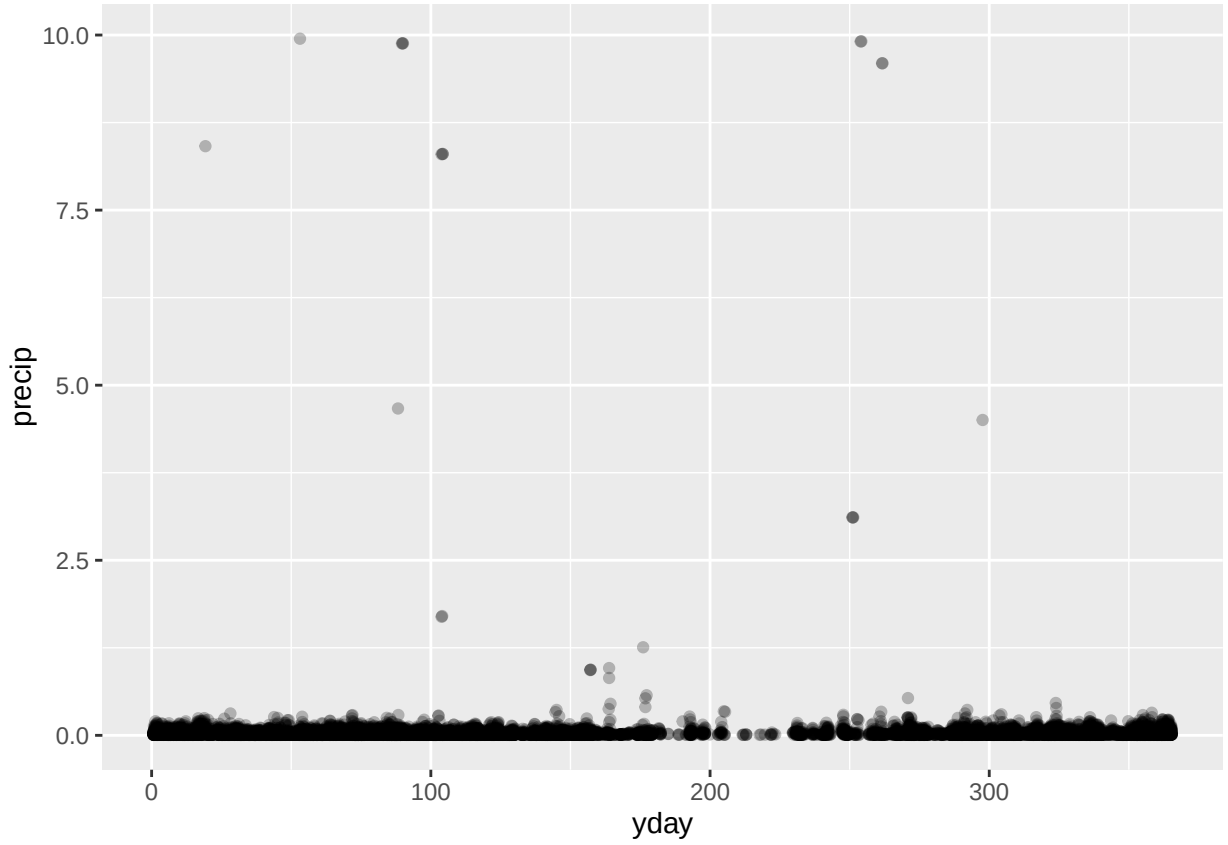
```
corv$month <- month(corv$date)
corv$year <- year(corv$date)
corv$yday <- yday(corv$date)

head(corv)
```

```
##              datetime  datePOSIX hour temp precip    conditions       date
## 1 2005-01-01 00:15:00 2005-01-01    0 42.8     NA      Overcast 2005-01-01
## 2 2005-01-01 00:35:00 2005-01-01    0 42.8     NA Mostly Cloudy 2005-01-01
## 3 2005-01-01 00:55:00 2005-01-01    0 42.8     NA      Overcast 2005-01-01
## 4 2005-01-01 01:15:00 2005-01-01    1 42.8     NA Mostly Cloudy 2005-01-01
## 5 2005-01-01 01:35:00 2005-01-01    1 42.8     NA Mostly Cloudy 2005-01-01
## 6 2005-01-01 01:55:00 2005-01-01    1 42.8     NA      Overcast 2005-01-01
##   month year yday
```

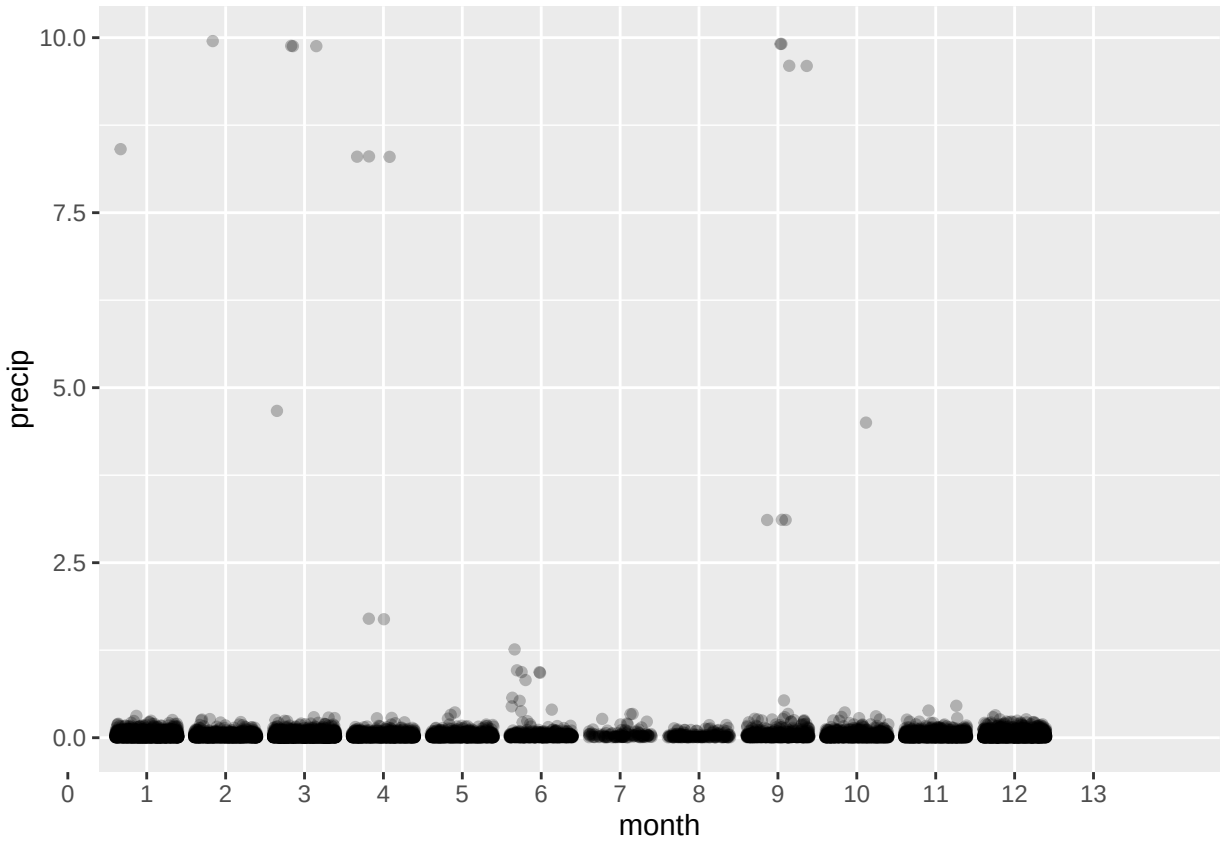
```
## 1      1 2005     1
## 2      1 2005     1
## 3      1 2005     1
## 4      1 2005     1
## 5      1 2005     1
## 6      1 2005     1
```

```
qplot(yday, precip, data = filter(corv, !is.na(precip)), geom = "jitter", alpha = I(1/4))
```



Q2.4. Which days of the year are rainy?

```
qplot(month, precip, data = filter(corv, !is.na(precip)), geom = "jitter", alpha = I(1/4)) + scale_x_dis
```


10.0
7.5
5.0
2.5
0.0
precip
0
1
2
3
4
5
6
7
8
9
10
11
12
13
month