Assignment 3

Bailey Rosato

2023-10-24

Chapter 11

Question 1

1. For the following regular expression, explain in words what it matches on. Then add test strings to demonstrate that it in fact does match on the pattern you claim it does. Make sure that your test set of strings has several examples that match as well as several that do not. *If you copy the Rmarkdown code for these exercises directly from my source pages, make sure to remove the* `eval=FALSE` *from the R-chunk headers.*
   a) This regular expression matches: any string with the letter 'a' exactly.

```
strings <- c("cba")
string <- "abc"
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, 'a') )
```

```
##   string result
## 1    cba   TRUE
```

b) This regular expression matches: any string with the letters 'ab' exactly.

```
# This regular expression matches:  Insert your answer here...
strings <- c("ab")
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, 'ab') )
```

c) This regular expression matches: an string with only a or b.

```
strings <- c("")
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '[ab]') )
```

```
##   string result
## 1         FALSE
```

d) This regular expression matches: only a or b at the beginning of the input string.

```
strings <- c()
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '^[ab]') )
```

```
## [1] result
## <0 rows> (or 0-length row.names)
```

e) This regular expression matches: any digit before white space between

only a or A.

```
        strings <- c("a3 A")
        data.frame( string = strings ) %>%
          mutate( result = str_detect(string, '\\d+\\s[aA]') )
```

```
##   string result
## 1   a3 A   TRUE
```

f) This regular expression matches: any string that starts with a digit before white space between only a and A with any repetition of 0 or above.

```
        strings <- c("a 3A a 5A")
        data.frame( string = strings ) %>%
          mutate( result = str_detect(string, '\\d+\\s*[aA]') )
```

```
##      string result
## 1 a 3A a 5A   TRUE
```

g) This regular expression matches: any string with any character with any amount of repetitions as a pattern.

```
        strings <- c("a5 87", " ")
        data.frame( string = strings ) %>%
          mutate( result = str_detect(string, '.*') )
```

```
##   string result
## 1  a5 87   TRUE
## 2          TRUE
```

h) This regular expression matches: any string that begins with an alphanumeric character and matches exactly two word characters followed by "bar" exactly.

```
        strings <- c("avbar", "abar", "23bar", "%@bar")
        data.frame( string = strings ) %>%
          mutate( result = str_detect(string, '^\\w{2}bar') )
```

```
##   string result
## 1  avbar   TRUE
## 2   abar  FALSE
## 3  23bar   TRUE
## 4  %@bar  FALSE
```

i) This regular expression matches: any string that matches the characters "foo.bar" exactly with the dot character or any string that begins with an alphanumeric character and matches exactly two word characters followed by "bar" exactly.

```
        strings <- c("foo.bar", "foobar", "avbar", "fobar")
        data.frame( string = strings ) %>%
          mutate( result = str_detect(string, '(foo\\.bar)|(^\\w{2}bar)') )
```

```
##    string result
## 1 foo.bar   TRUE
## 2  foobar  FALSE
## 3   avbar   TRUE
## 4   fobar   TRUE
```

Question 2

2. The following file names were used in a camera trap study. The S number represents the site, P is the plot within a site, C is the camera number within the plot, the first string of numbers is the YearMonthDay and the second string of numbers is the HourMinuteSecond.

```
file.names <- c( 'S123.P2.C10_20120621_213422.jpg',
                 'S10.P1.C1_20120622_050148.jpg',
                'S187.P2.C2_20120702_023501.jpg')

df <- data.frame(
     file_info = file.names ) %>%
     cbind( str_split_fixed(.$file_info, pattern='[\\.\\_]', n=6)) %>%
     rename( Site=`1`, Plot=`2`, Camera=`3`, Date = `4`, Time=`5`) %>%
     mutate(Year = str_sub(Date, 1, 4),
            Month = str_sub(Date, 5, 6),
            Day = str_sub(Date, 7,8)) %>%
     mutate(Hour = str_sub(Time, 1,2),
            Minute = str_sub(Time, 3,4),
            Second = str_sub(Time, 5,6)) %>%
     select(-Date) %>%
     select(-Time) %>%
     select(-file_info) %>%
     select(-`6`)

# save multiple objects and then split them up separately into the data frame
df
```

```
##   Site Plot Camera Year Month Day Hour Minute Second
## 1 S123   P2    C10 2012    06  21   21     34     22
## 2  S10   P1     C1 2012    06  22   05     01     48
## 3 S187   P2     C2 2012    07  02   02     35     01
```

Produce a data frame with columns corresponding to the `site`, `plot`, `camera`, `year`, `month`, `day`, `hour`, `minute`, and `second` for these three file names. So we want to produce code that will create the data frame:

Site	Plot	Camera	Year	Month	Day	Hour	Minute	Second
S123	P2	C10	2012	06	21	21	34	22
S10	P1	C1	2012	06	22	05	01	48
S187	P2	C2	2012	07	02	02	35	01

Question 3

3. The full text from Lincoln's Gettysburg Address is given below. Calculate the mean word length *Note: consider 'battle-field' as one word with 11 letters*).

```
Gettysburg <- 'Four score and seven years ago our fathers brought forth on this
continent, a new nation, conceived in Liberty, and dedicated to the proposition
that all men are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any
nation so conceived and so dedicated, can long endure. We are met on a great
battle-field of that war. We have come to dedicate a portion of that field, as
a final resting place for those who here gave their lives that that nation might
live. It is altogether fitting and proper that we should do this.
```

```
But, in a larger sense, we can not dedicate -- we can not consecrate -- we can
not hallow -- this ground. The brave men, living and dead, who struggled here,
have consecrated it, far above our poor power to add or detract. The world will
little note, nor long remember what we say here, but it can never forget what
they did here. It is for us the living, rather, to be dedicated here to the
unfinished work which they who fought here have thus far so nobly advanced. It
is rather for us to be here dedicated to the great task remaining before us --
that from these honored dead we take increased devotion to that cause for which
they gave the last full measure of devotion -- that we here highly resolve that
these dead shall not have died in vain -- that this nation, under God, shall
have a new birth of freedom -- and that government of the people, by the people,
for the people, shall not perish from the earth.'

df <- data.frame( string = Gettysburg ) %>%
           mutate( result = str_split(string, '[\\.\\s\\,\\--]+'))

# df$result

wordmean<-str_length(df$result[[1]])

numStrings<-(length(df$result[[1]]))

wordavg<-sum(wordmean)/numStrings

wordavg
```

```
## [1] 4.208791
```

Chapter 12

Question 1

1. Convert the following to date or date/time objects.
   a) September 13, 2010.

```
mdy('September 13, 2010.')
```

```
## [1] "2010-09-13"
```

b) Sept 13, 2010.

```
#data_object <- as.Date('Sept 13, 2010.', format='%d%b%Y')

#data_object
```

This does not work because the abbreviation is not correct.

c) Sep 13, 2010.

```
mdy('Sep 13, 2010.')
```

```
## [1] "2010-09-13"
```

d) S 13, 2010. Comment on the month abbreviation needs.

```
# mdy('S 13, 2010.')
```

The function does not work when the month is abbreviated to only one letter. It requires the whole month name or a 3 letter abbreviation.

e) 07-Dec-1941.

```
dmy('07-Dec-1941.')
```

```
## [1] "1941-12-07"
```

f) 1-5-1998. Comment on why you might be wrong.

```
mdy('1-5-1998')
```

```
## [1] "1998-01-05"
```

This may be wrong because the 1 or 5 can be referring to the month or the day of the month and there is no way to tell from the formatting.

g) 21-5-1998. Comment on why you know you are correct.

```
dmy('21-5-1998')
```

```
## [1] "1998-05-21"
```

This is correct because the number 21 indicates the day and not the month because we know there are only twelve months.

h) 2020-May-5 10:30 am

```
ymd_hm('2020-May-5 10:30 am')
```

```
## [1] "2020-05-05 10:30:00 UTC"
```

i) 2020-May-5 10:30 am PDT (ex Seattle)

```
ymd_hm('2020-May-5 10:30 am', tz='US/Pacific')
```

```
## [1] "2020-05-05 10:30:00 PDT"
```

j) 2020-May-5 10:30 am AST (ex Puerto Rico)

```
ymd_hm('2020-May-5 10:30 am', tz='America/Antigua')
```

```
## [1] "2020-05-05 10:30:00 AST"
```

Question 2

2. Using just your date of birth (ex Sep 7, 1998) and today's date calculate the following *Write your code in a manner that the code will work on any date after you were born.*:
   a) Calculate the date of your 64th birthday.

```
birthdate <- mdy('Nov 21, 2001')

birthday64 = birthdate + years(64)

birthday64
```

```
## [1] "2065-11-21"
```

b) Calculate your current age (in years). _Hint: Check your age is calculated correctly if your birthday was yesterday and if it were tomorrow!_

```
birthdate <- ymd('2001 Nov 21')

today <- Sys.Date()

currentAge <- interval(birthdate, today)
```

```r
currentAge <- as.numeric(as.duration(currentAge), 'years')

currentAge
```

```
## [1] 21.92197
```

When calculated the date of birth was yesterday, the age is 0 and it is also 0 when the date of birth is tomorrow. d) Using your result in part (b), calculate the date of your next birthday.

```r
nextAge <- currentAge + 1

next_year <- as.numeric(format(birthdate, "%Y")) + nextAge

nextbirthday <- make_date(year=next_year, month=11, day=21)

nextbirthday
```

```
## [1] "2023-11-21"
```

e) The number of _days_ until your next birthday.

```r
daysTillBday <- as.numeric(nextbirthday - today)

daysTillBday
```

```
## [1] 28
```

f) The number of _months_ and _days_ until your next birthday.

```r
monthsTillBday <- floor(daysTillBday / 30)

days <- daysTillBday - monthsTillBday * 30

cat("There are", monthsTillBday,"months and",days,"days until my next birthday")
```

```
## There are 0 months and 28 days until my next birthday
```

Question 3

3. Suppose you have arranged for a phone call to be at 3 pm on May 8, 2015 at Arizona time. However, the recipient will be in Auckland, NZ. What time will it be there?

```r
AZTime <- ymd_hms('2015-05-08 15:00:00', tz='America/Phoenix')

aucklandTime <- with_tz(AZTime, tz='Pacific/Auckland')

aucklandTime
```

```
## [1] "2015-05-09 10:00:00 NZST"
```

Question 5

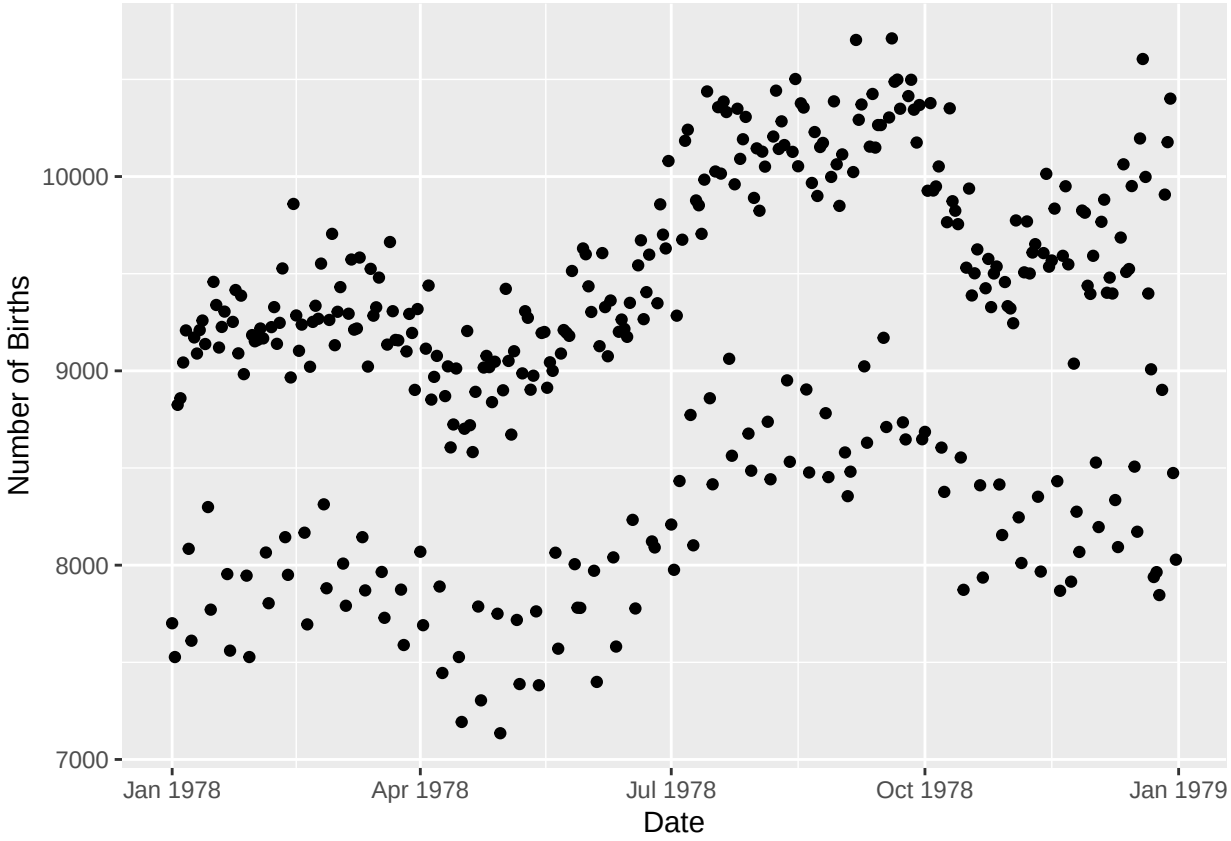
5. It turns out there is some interesting periodicity regarding the number of births on particular days of the year.
   a. Using the `mosaicData` package, load the data set `Births78` which records the number of children born on each day in the United States in 1978. Because this problem is intended to show how to calculate the information using the `date`, remove all the columns *except* `date` and `births`.

```
data("Births78")

dataBirths <- Births78[, c("date", "births")]

head(dataBirths)
```

```
##         date births
## 1 1978-01-01   7701
## 2 1978-01-02   7527
## 3 1978-01-03   8825
## 4 1978-01-04   8859
## 5 1978-01-05   9043
## 6 1978-01-06   9208
```

b. Graph the number of `births` vs the `date` with date on the x-axis. What stands out to you? Why do y

```
ggplot(dataBirths, aes(x=date, y=births)) + geom_point() +
        labs(x="Date", y="Number of Births")
```



There is a trend of higher births in the months July through October and there is a lower amount of births in the months of April through July. This can be caused by seasonal differences in weather.

c. To test your assumption, we need to figure out the what day of the week each observation is. Use `dpl

```
dataBirths <- dataBirths %>% mutate(dow = wday(date, label=TRUE, abbr=FALSE))

head(dataBirths)
```

```
##         date births       dow
## 1 1978-01-01   7701    Sunday
```

```
## 2 1978-01-02   7527    Monday
## 3 1978-01-03   8825   Tuesday
## 4 1978-01-04   8859 Wednesday
## 5 1978-01-05   9043  Thursday
## 6 1978-01-06   9208    Friday
```

d. Plot the data with the point color being determined by the day of the week variable.

```
ggplot(dataBirths, aes(x=date, y=births, color=dow)) + geom_point() +
        labs(x="Date", y="Number of Births", Color="Day of Week")
```