445_Assignment_3

Tucker Harris

2023-10-26

Chapter 11 question 1a

This detects when a string in the list of strings contains the character 'a'. This can be seen when plugging in three words with the letter a, and one without.

```
strings <- c("Alphabet", "Airplane","Frog", "Zebra" )
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, 'a') )
```

```
##      string result
## 1 Alphabet   TRUE
## 2 Airplane   TRUE
## 3     Frog  FALSE
## 4    Zebra   TRUE
```

b.)

This detect will determine if a given string contains the substring 'ab' in the word.This is case sensitive.

```
strings <- c("Falling", "into", "The", "abyss")
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, 'ab') )
```

```
##     string result
## 1 Falling  FALSE
## 2    into  FALSE
## 3     The  FALSE
## 4   abyss   TRUE
```

c.) This will search for the string containing an 'a' or a 'b', the brackets do not need to be a part of the string.

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

```
##   string result
## 1   [ab]   TRUE
## 2  climb   TRUE
## 3   cook  FALSE
## 4   beef   TRUE
## 5  alone   TRUE
```

d.)

This expression determines if it starts with either an 'a' or a 'b'. This is still case sensitive, and the word can start with either or both characters.

```r
strings <- c("aa", "B", "A", "bean", "Clip")
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '^[ab]') )
```

```
##   string result
## 1     aa   TRUE
## 2      B  FALSE
## 3      A  FALSE
## 4   bean   TRUE
## 5   Clip  FALSE
```

e.)

This should match any digit directly followed by white space with the next character being an 'a' or an 'A'

```r
strings <- c("1 a", "5 khaki", "16 alone", "18a", "Scoring A 95", "88 was the class average")
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '\\d+\\s[aA]') )
```

```
##                     string result
## 1                      1 a   TRUE
## 2                  5 khaki  FALSE
## 3                 16 alone   TRUE
## 4                      18a  FALSE
## 5             Scoring A 95  FALSE
## 6 88 was the class average  FALSE
```

f.) This string matches a digit directly followed by either the character 'a' or 'A' whether or not there is white space in between the digit and the character.

```r
strings <- c("6a", "6      ba", "4        ab", "778Alkaline")
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '\\d+\\s*[aA]') )
```

```
##        string result
## 1          6a   TRUE
## 2   6      ba  FALSE
## 3 4        ab   TRUE
## 4 778Alkaline   TRUE
```

g.) This regular expression matches: a string with any character

```r
strings <- c("", "16", "This", 16)
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '.*') )
```

```
##   string result
## 1          TRUE
## 2     16   TRUE
## 3   This   TRUE
## 4     16   TRUE
```

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

```
##    string result
## 1   ffbar   TRUE
## 2 foo.bar   TRUE
```

```
## 3    %%bar   FALSE
## 4   foobar   FALSE
```

Chapter 11 Question 2

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

x1 <- str_replace_all(file.names, pattern = '_', replacement = '.')

x2 <- str_split_fixed( x1, pattern = '\\.', n = 6 )

year <- str_sub(x2[,4], start = 1, end = 4 )

month <- str_sub(x2[,4], start = 5, end = 6)

day <- str_sub(x2[,4], start = 7, end = 8)

hour <- str_sub(x2[,5], start = 1, end = 2)

minute <- str_sub(x2[,5], start = 3, end = 4)

second <- str_sub(x2[,5], start = 5, end = 6)

Site <- x2[,1]

Plot <- x2[,2]

Camera <- x2[,3]

data.frame(
  site = Site,
  plot = Plot,
  camera = Camera,
  year = year,
  month = month,
  day = day,
  hour = hour,
  minute = minute,
  second = second


)
```

```
##   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
```

Chapter 11 Question 3

```r
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.'


Gettysburg <- str_remove_all(Gettysburg, '\\.')

Gettysburg <- str_remove_all(Gettysburg, ',')

Gettysburg <- str_remove_all(Gettysburg, '-')

Gettysburg <- str_trim(Gettysburg)

splitBurg <- str_split(Gettysburg, '\\s+')

# splitBurg[[1]]

speechLength <- length(splitBurg[[1]])

speechChars <- sum(str_length(splitBurg[[1]]))

# The average word length is now the sum of the length divided by the number of words.

speechChars / speechLength
```

[1] 4.239852

For some reason, I was unable to compile with part h and part i. I had those working, but they would not
compile for some reason. I apologize for that.

Chapter 12 question 1

```
TimeObject1 <- 'September 13, 2010'

mdy(TimeObject1)
```

[1] "2010-09-13"

```r
TimeObject2 <- 'Sept 13, 2010'

as.Date(TimeObject2, format = '%b %d, %Y')
```

```
## [1] NA
```
```r
# This one comes out as NA, I believe it cannot read it in the format it was given. However, it is bett

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

```
## [1] "2010-09-13"
```
```r
#mdy('S 13, 2010')
# No formats found. Using S is not sufficient and R studio cannot interpret it

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

```
## [1] "1941-12-07"
```
```r
mdy('1-5-1998') # This one could be wrong, because it could be referring to either January 5th or May 1
```

```
## [1] "1998-01-05"
```
```r
dmy('21-5-1998') # This one is correct, as there is no 21 month, so it can only be reffering to May 21
```

```
## [1] "1998-05-21"
```
```r
ymd_hm('2020-May-5 10:30 am')
```

```
## [1] "2020-05-05 10:30:00 UTC"
```
```r
ymd_hm('2020-May-5 10:30 am', tz = 'US/Pacific')
```

```
## [1] "2020-05-05 10:30:00 PDT"
```
```r
ymd_hm('2020-May-5 10:30 am', tz = 'America/Puerto_Rico')
```

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

Chapter 12 Question 2a

```r
Birthday <- mdy('03-21-2003')

Today <- today()

SixtyFive <- Birthday + years(65)

TimeTilSixtyFive <- interval(Today, SixtyFive)

as.period(TimeTilSixtyFive)
```

```
## [1] "44y 4m 24d 0H 0M 0S"
```

Question 2 b

```r
Age <- interval(Birthday, Today)

Age <- as.period(Age)
```

```
year(Age)
```

```
## [1] 20
```

Question 2 c

```
nextBday <- Birthday + years(21)

nextBday
```

```
## [1] "2024-03-21"
```

Question 2 d

```
WaitTime <- interval(Today, nextBday)

WaitTime <- as.period(WaitTime, unit = 'days')

WaitTime
```

```
## [1] "147d 0H 0M 0S"
```

Chapter 12 Question 2e

```
WaitTime <- interval(Today, nextBday)

WaitTime <- as.period(WaitTime)

WaitTime
```

```
## [1] "4m 24d 0H 0M 0S"
```

Chapter 12 Question 3

```
Call <- ymd_hm('2015-May-8 3:00 pm', tz = 'US/Arizona')

with_tz(Call, 'Pacific/Auckland')
```

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

Chapter 12 Question 5 a

```
library(mosaicData)


data("Births78")

Births78 <- Births78 %>% select(date,births)

head(Births78)
```
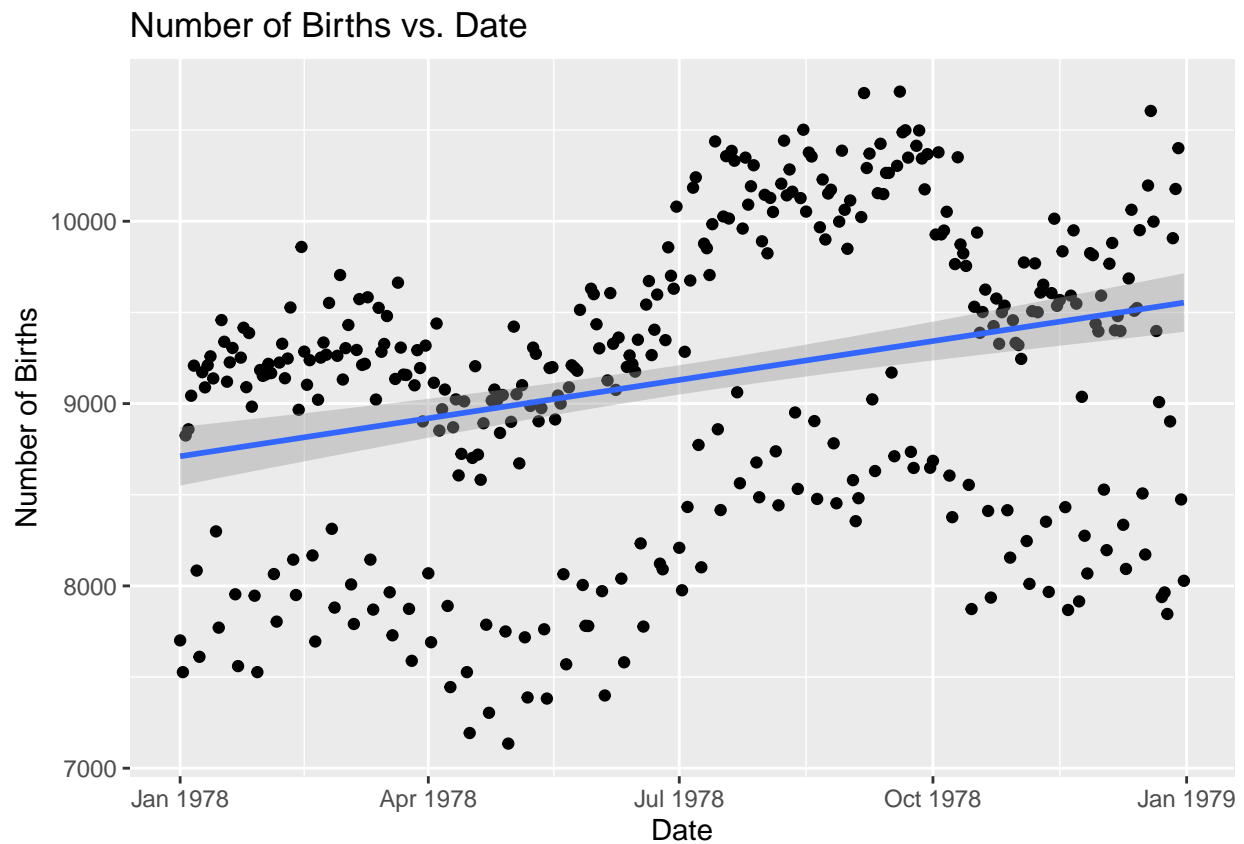
```
##         date births
```

```
## 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
```

Chapter 12 question 5b

```
ggplot(Births78, aes(x= date, y = births)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs( title = "Number of Births vs. Date",
        x = "Date",
        y = "Number of Births")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


There appears to be a weak positive linear relationship between number of births and time. Over time, the number of births seems to increase, but the model does not appear that it would have a very high R squared value.

Chapter 12 Question 5c

```
Births78 <- Births78 %>%
  mutate(dow = wday(date, label = TRUE, abbr = FALSE, ))

head(Births78)
```

```
##          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
```

Chapter 12 Question 5d

```
ggplot(Births78, aes(x = date, y = births, color = dow)) +
  geom_point() +
  geom_smooth( method = 'lm' ) +
  labs(title = "Number of Births vs. Date",
       x = "Date",
       y = "Number of Births")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```