Yurong Jiang

Homework04

```
#1
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(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
epl <- read.csv("epl.csv")
starbucks <- read.csv('starbucks_US.csv')

epl %>%
  filter( ( (HomeTeam == 'Liverpool' & Home_Win == 1) |
              (AwayTeam == 'Liverpool' & Away_Win == 1) ) &
            (ymd(Match_Date) >= ymd('2001-01-01') & ymd(Match_Date) < ymd('2011-01-01'))) %>%
  summarise(n())
```

```
##   n()
## 1 200
```

```
#2
epl %>%
group_by(year(ymd(Match_Date))) %>%
summarise(num = n()) %>%
arrange(desc(num)) %>%
head(5)
```

```
## # A tibble: 5 x 2
##   `year(ymd(Match_Date))`   num
##                     <dbl> <int>
## 1                    1994   456
## 2                    1995   426
## 3                    2017   401
## 4                    2006   394
## 5                    2004   392
```

```r
#3.
epl %>%
filter(year(ymd(Match_Date)) == 2015) %>%
group_by(month(ymd(Match_Date))) %>%
summarise(num = n())
```

```
## # A tibble: 10 x 2
##    `month(ymd(Match_Date))`   num
##                       <dbl> <int>
##  1                        1    38
##  2                        2    38
##  3                        3    33
##  4                        4    40
##  5                        5    41
##  6                        8    40
##  7                        9    30
##  8                       10    37
##  9                       11    33
## 10                       12    50
```

```r
#4
epl %>%
  filter(year(Match_Date) == 2017) %>%
  group_by(month(Match_Date),day(Match_Date)) %>%
  summarise(max = n()) %>%
  arrange(-max) %>%
  head(1)
```

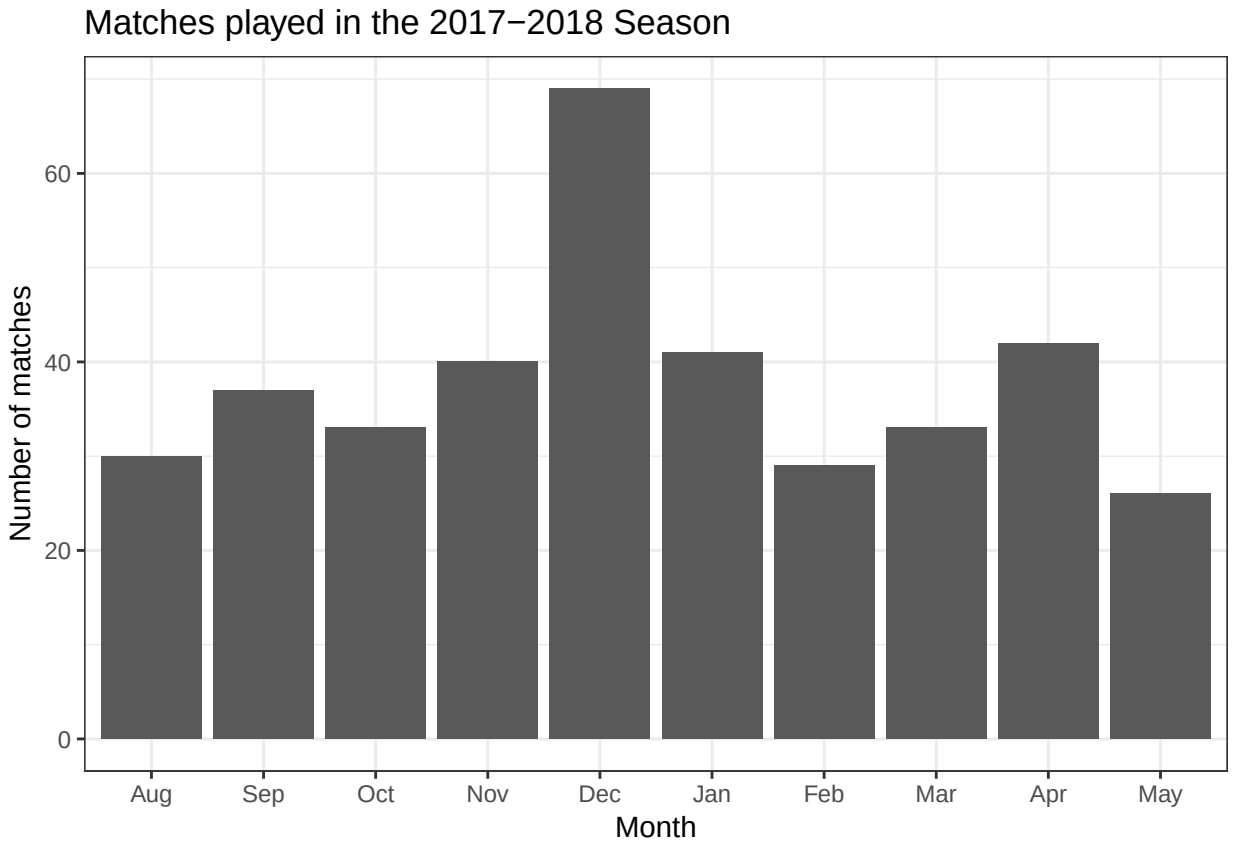
```
## # A tibble: 1 x 3
## # Groups:   month(Match_Date) [1]
##   `month(Match_Date)` `day(Match_Date)`   max
##                 <dbl>             <int> <int>
## 1                   5                21    10
```

```r
#5
epl %>%
filter(Season == "2017-18") %>%
group_by(wday=wday(ymd(Match_Date), label = T)) %>%
summarise(count = n())
```

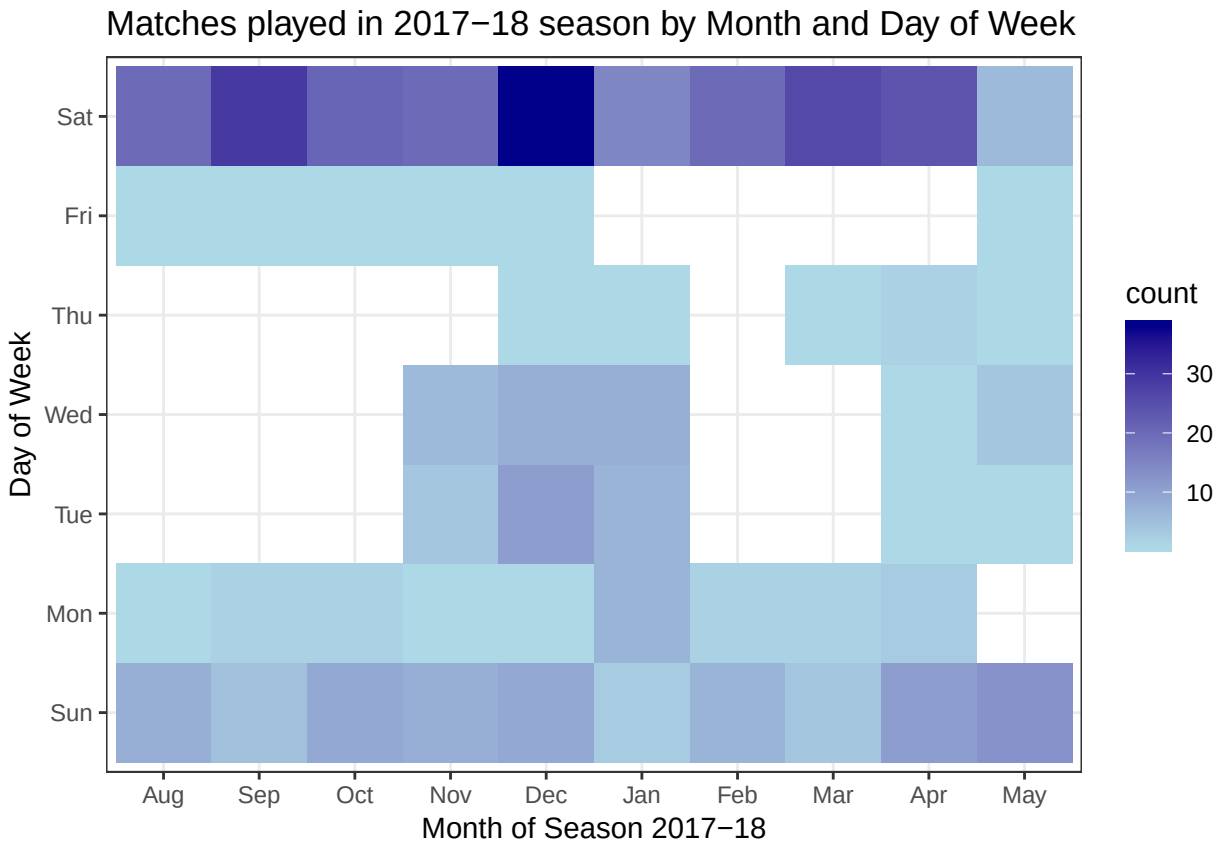
```
## # A tibble: 7 x 2
##   wday  count
##   <ord> <int>
## 1 Sun      77
## 2 Mon      21
## 3 Tue      24
## 4 Wed      27
## 5 Thu       6
## 6 Fri       6
## 7 Sat     219
```

```r
#6
epl %>%
  mutate(month = as.factor(month(ymd(Match_Date), label= T, abbr = T))) %>%
  filter(Season == '2017-18') %>%
  mutate(month = factor(month, levels = c('Aug', 'Sep', 'Oct', 'Nov', 'Dec', 'Jan', 'Feb', 'Mar', 'Apr'
```

```
  ggplot(aes(x = month))+
  geom_bar() +
  xlab('Month') +
  ylab('Number of matches') +
  ggtitle('Matches played in the 2017-2018 Season') +
  theme_bw()
```

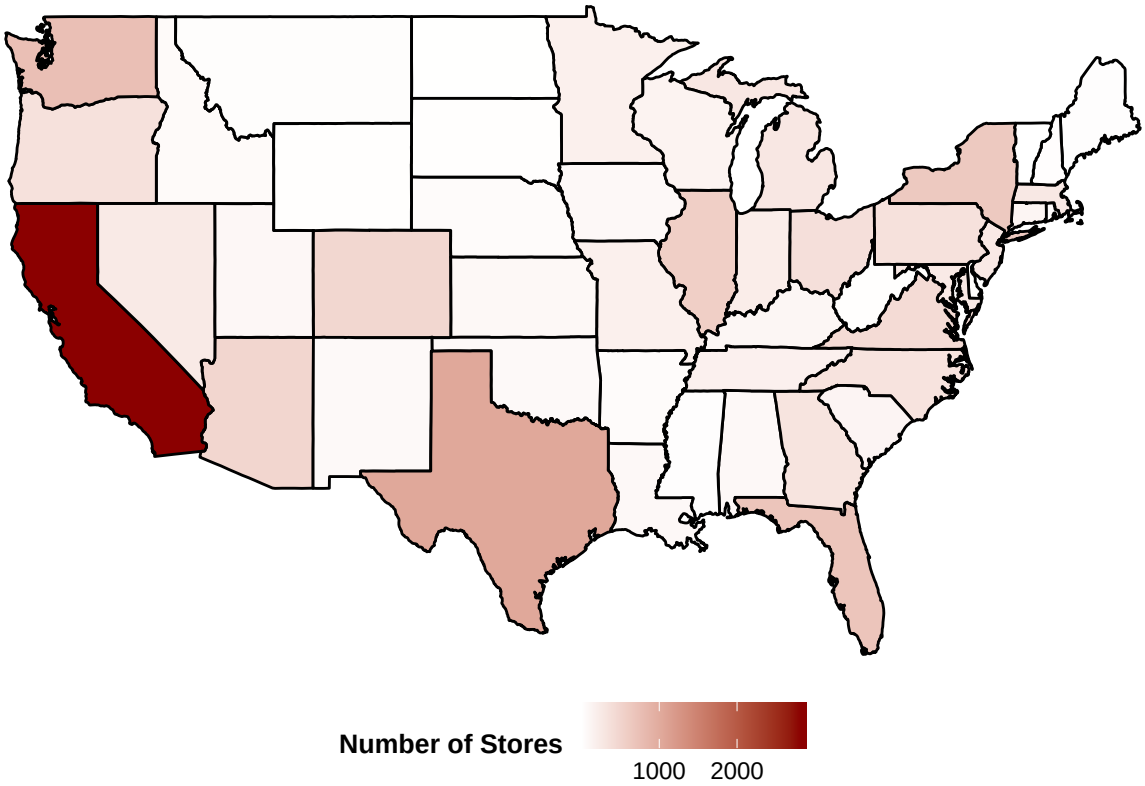



```
#7

epl %>%
  mutate(wday = wday(ymd(Match_Date), label = T, abbr = T)) %>%
  mutate(month = as.factor(month(ymd(Match_Date),label = T, abbr = T))) %>%
  mutate(month =
           factor(month, levels =c('Aug', 'Sep', 'Oct', 'Nov', 'Dec', 'Jan', 'Feb', 'Mar', 'Apr', 'May'
  filter(Season == '2017-18') %>%
  group_by(month, wday) %>%
  summarise(count = n()) %>%
  ggplot(aes(x = month,
             y = wday,
             fill = count)) +
  geom_tile() +
  scale_fill_gradient(low = 'lightblue', high = 'darkblue') +
  theme_bw() +
  xlab('Month of Season 2017-18') +
  ylab('Day of Week') +
  ggtitle('Matches played in 2017-18 season by Month and Day of Week')
```




```
#8
state_usa = map_data("state")
starbucks$State = tolower(starbucks$State)
data1 = starbucks %>%
  group_by(State) %>%
  summarise(star_num = n())
nofstar = left_join(state_usa, data1,
                          by = c('region' = 'State'),
                          merge(state_usa, starbucks))


ggplot(nofstar, aes(x = long, y = lat,fill = star_num,group = group)) +
  geom_polygon(color = 'black') +
  ggtitle('Starbucks Franchise across United States') +
  labs( fill = 'Number of Stores')+
  theme_void() +
  theme(legend.position = 'bottom',
        plot.title = element_text(face = 'bold', size = 16),
        legend.title = element_text(face = 'bold', size = 10))+
    scale_fill_gradient(low = 'white',
                        high = 'darkred')
```

Starbucks Franchise across United States



9

Key point: 1.We experience cognitive load anytime we take in information. Cognitive load can be thought of as the mental effort that??s required to learn new information. 2.One culprit that can contribute to excessive or extraneous cognitive load is something I refer to simply as clutter. These are visual elements that take up space but don??t increase understanding. 3.Gestalt principles of visual perception??proximity, similarity, enclosure, closure, continuity, and connection. 4.Without other visual cues, your audience will typically start at the top left of the page or screen and will move their eyes in a ??z?? shape (or multiple ??z?? shapes, depending on the layout) across the page or screen as they take in information. Because of this, when it comes to tables and graphs, I like to upper-left-most justify the text (title, axis titles, legend). This means the audience will hit the details that tell them how to read the table or graph before they get to the data itself.White space in visual communication is as important as pauses in public speaking. 5.If there is something really important we want our audience to know or see (the hawk), we 6. Six major changes to reduce clutter:(1). Remove chart border (2)Remove gridlines (3)Remove data markers(4)Clean up axis labels(5)Label data directly(6)Leverage consistent color 7.Any time you put information in front of your audience, you are creating cognitive load and asking them to use their brain power to process that information. Visual clutter creates excessive cognitive load that can hinder the transmission of our message. The Gestalt Principles of Visual Perception can help you understand how your audience sees and allow you to identify and remove unnecessary visual elements. Leverage alignment of elements and maintain white space to help make the interpretation of your visuals a more comfortable experience for your audience. Use contrast strategically. Clutter is your enemy.