Week 12 In Class Lecture

Peyton Hall

03/28/2024

```
rm(list=ls())
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# library(ggplot2)
library(nycflights13)
# install.packages("plotly")
library(plotly)
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:ggplot2':
##
##     last_plot
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following object is masked from 'package:graphics':
##
##     layout
```

```
# install.packages("DT")
library(DT)
# install.packages("stringr")
library(stringr)
# install.packages("tm")
library(tm)
```

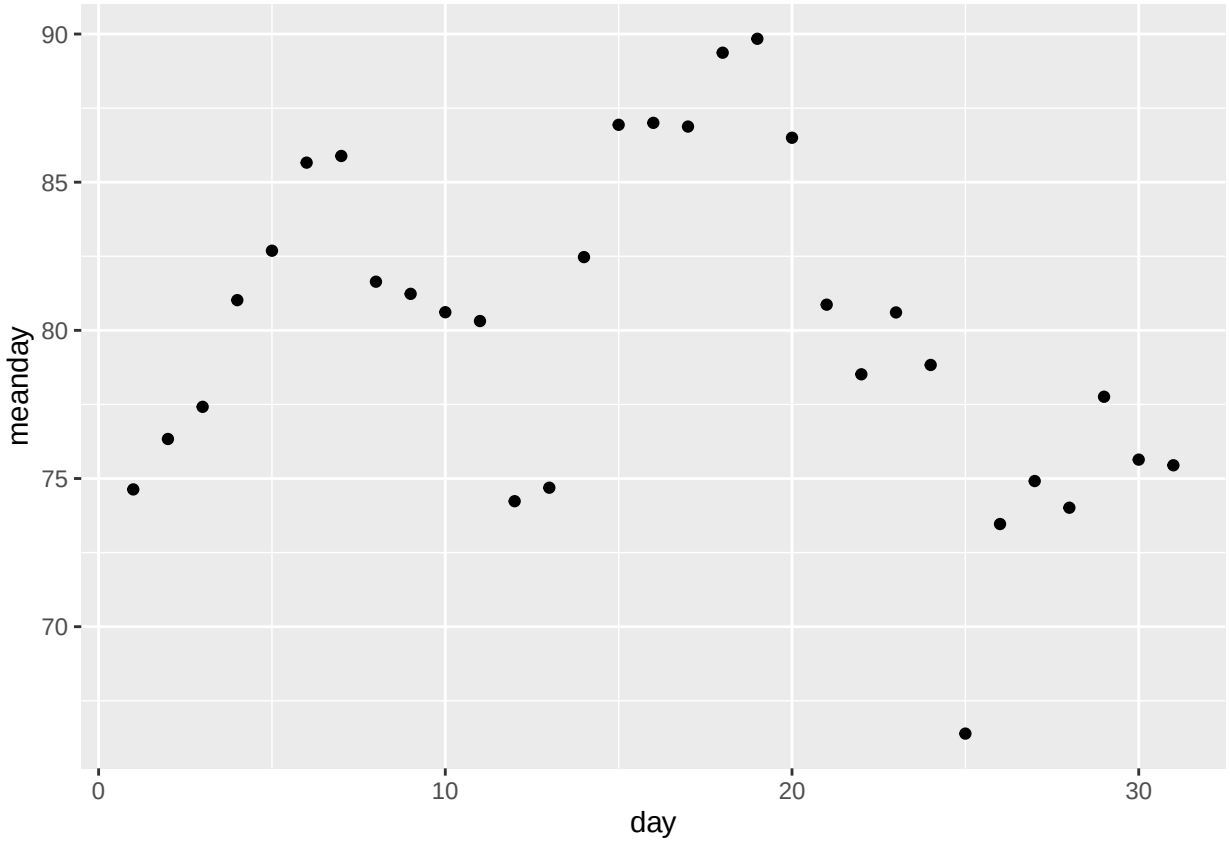
```
## Loading required package: NLP
##
## Attaching package: 'NLP'
##
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
# install.packages("wordcloud")
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

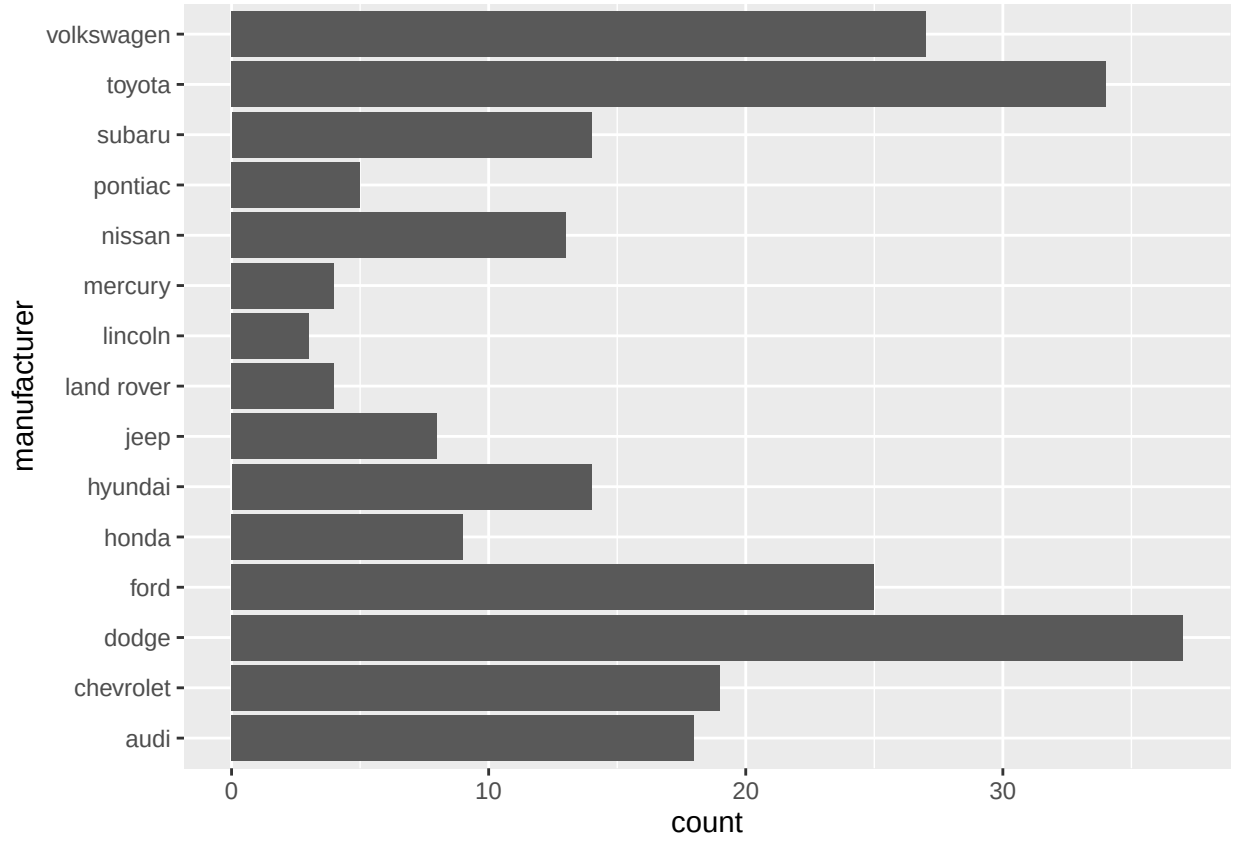
ggplotly()

```
# View(weather)
graph1 <- weather %>%
  filter(month==7)%>%
  group_by(day)%>%
  summarise(meanday=mean(temp)) %>%
  ggplot(aes(x=day, y=meanday))+geom_point()
graph1
```



```
# here is a more interactive plot:
graph2 <- ggplotly(graph1)
# graph2
```

```
Activity1 <- ggplot(data = mpg,aes(y=manufacturer))+geom_bar()
Activity1
```




```
Activity01 <- ggplotly(Activity1)
# Activity 01
```

Example

```
employee <- data.frame(
  Name = c("John", "Kate", "Tom"),
  Title = c("Manager", "Student", "Professor"),
  Salary = c(70000, 30000, 60000)
)
employee
```

```
##   Name     Title Salary
## 1 John   Manager  70000
```

```
## 2 Kate   Student  30000
## 3  Tom Professor  60000
```

```
emplotee1 <- datatable(employee)
# emplotee1
```

```
Activity2 <- data.frame(Student = c("Jack", "Mike", "Kate", "Mary"),
  Midterm = c(89, 76, 76, 90),
  Final = c(91, 72, 81, 92)
)
Activity02 <- datatable(Activity2)
# Activity02
```

```
x <- c ("Anderson", "Jackson", "Peterson", "Green")
x[grep("son", x)] # it is case sensitive
```

```
## [1] "Anderson" "Jackson"  "Peterson"
```

```
substr(x,start=1,stop=4)
```

```
## [1] "Ande" "Jack" "Pete" "Gree"
```

```
# substitution
gsub("son","",x) # search for "son" and replace with "" (blank).
```

```
## [1] "Ander" "Jack"  "Peter" "Green"
```

```
str_replace_all(x,"son", "")
```

```
## [1] "Ander" "Jack"  "Peter" "Green"
```

```
# Analyze the text in terms of word frequencies
#   Usually analyze the frequencies of words and generate word cloud
#   Example:
# Use the text of the MPR news of Vice President Kamala Harris visiting the Twin
# Cities and the roundtable event at Metro State University in October 2022 to
# do text mining and generate a word cloud.

# Step 1: Data Importing and Cleaning
# Import the data into R and clean it to remove any irrelevant information, special
# characters, etc.
Mytext1<-readLines("https://www.mprnews.org/story/2022/10/22/theres-so-much-at-stake-vp-harris-makes-tw
```

```
## Warning in
## readLines("https://www.mprnews.org/story/2022/10/22/theres-so-much-at-stake-vp-harris-makes-twin-cit
## incomplete final line found on
## 'https://www.mprnews.org/story/2022/10/22/theres-so-much-at-stake-vp-harris-makes-twin-cities-visit-
```

```r
# Creating a corpus (collection of text documents).
article <- Mytext1 %>%
  VectorSource() %>%
  Corpus() %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeWords,stopwords("English")) %>%
  tm_map(stripWhitespace) %>%
  TermDocumentMatrix() %>%
  as.matrix()
```

```
## Warning in tm_map.SimpleCorpus(., content_transformer(tolower)): transformation
## drops documents
```

```
## Warning in tm_map.SimpleCorpus(., removePunctuation): transformation drops
## documents
```

```
## Warning in tm_map.SimpleCorpus(., removeWords, stopwords("English")):
## transformation drops documents
```

```
## Warning in tm_map.SimpleCorpus(., stripWhitespace): transformation drops
## documents
```

```r
article2 <- as.data.frame(article)
article2$freq <- article2$`1`
```

```r
# Why one would do this:
# to answer the question of how frequent does a word show up in a text?
wordcloud(rownames(article2),article2$freq, min.freq = 7,
          colors = brewer.pal(8,"Dark2"))
```


```
Mytext1<-readLines("https://www.nytimes.com/2023/03/16/business/fed-regulation-svb.html")
```

```
## Warning in
## readLines("https://www.nytimes.com/2023/03/16/business/fed-regulation-svb.html"):
## incomplete final line found on
## 'https://www.nytimes.com/2023/03/16/business/fed-regulation-svb.html'
```

```
# Creating a corpus (collection of text documents).
article <- Mytext1 %>%
  VectorSource() %>%
  Corpus() %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeWords,stopwords("English")) %>%
  tm_map(stripWhitespace) %>%
  TermDocumentMatrix() %>%
  as.matrix()
```

```
## Warning in tm_map.SimpleCorpus(., content_transformer(tolower)): transformation
## drops documents
```

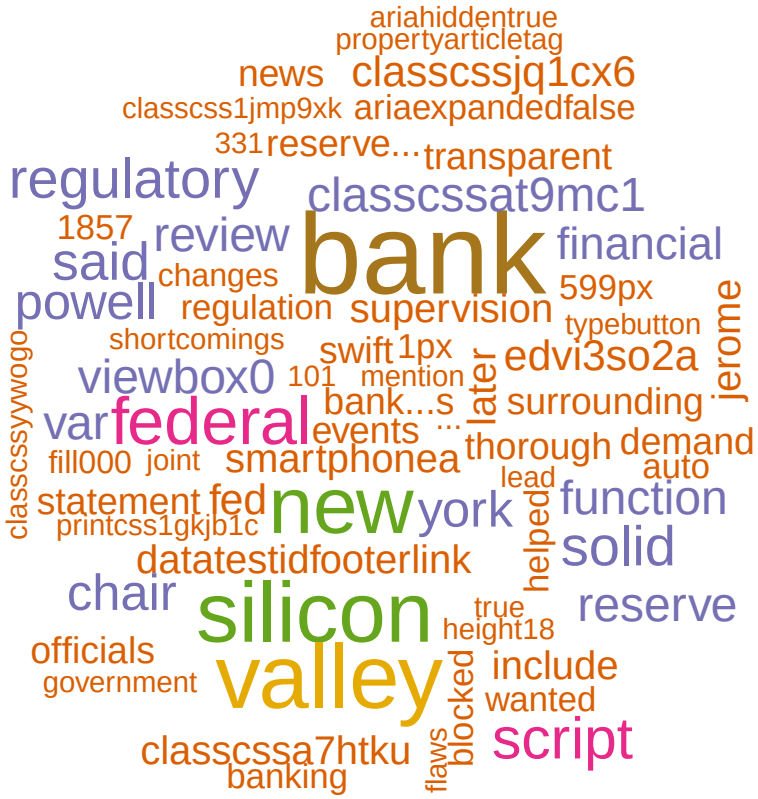
```
## Warning in tm_map.SimpleCorpus(., removePunctuation): transformation drops
## documents
```

```
## Warning in tm_map.SimpleCorpus(., removeWords, stopwords("English")):
## transformation drops documents

## Warning in tm_map.SimpleCorpus(., stripWhitespace): transformation drops
## documents
```

```
article1 <- sort(rowSums(article), decreasing = TRUE)
article2 <- as.data.frame(article1)

wordcloud(rownames(article2),article2$article1, min.freq = 9,
          colors = brewer.pal(8,"Dark2"))
```

```
## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'bank's' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'bank's' in 'mbcsToSbcs': dot substituted for <80>

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'bank's' in 'mbcsToSbcs': dot substituted for <99>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'bank's' in 'mbcsToSbcs': dot substituted for
## <e2>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'bank's' in 'mbcsToSbcs': dot substituted for
## <80>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'bank's' in 'mbcsToSbcs': dot substituted for
## <99>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font metrics unknown for Unicode character U+2019

## Warning in wordcloud(rownames(article2), article2$article1, min.freq = 9, :
## classcsswbbhzv could not be fit on page. It will not be plotted.

## Warning in wordcloud(rownames(article2), article2$article1, min.freq = 9, :
## classcss10t7hiaa could not be fit on page. It will not be plotted.

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'reserve"' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'reserve"' in 'mbcsToSbcs': dot substituted for <80>

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on
## 'reserve"' in 'mbcsToSbcs': dot substituted for <9d>
```

```
## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'reserve"' in 'mbcsToSbcs': dot substituted
## for <e2>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'reserve"' in 'mbcsToSbcs': dot substituted
## for <80>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on 'reserve"' in 'mbcsToSbcs': dot substituted
## for <9d>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font metrics unknown for Unicode character U+201d

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on '-' in
## 'mbcsToSbcs': dot substituted for <e2>

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on '-' in
## 'mbcsToSbcs': dot substituted for <80>

## Warning in strwidth(words[i], cex = size[i], ...): conversion failure on '-' in
## 'mbcsToSbcs': dot substituted for <94>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on '-' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on '-' in 'mbcsToSbcs': dot substituted for <80>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : conversion failure on '-' in 'mbcsToSbcs': dot substituted for <94>

## Warning in text.default(x1, y1, words[i], cex = size[i], offset = 0, srt =
## rotWord * : font metrics unknown for Unicode character U+2014

## Warning in wordcloud(rownames(article2), article2$article1, min.freq = 9, :
## datarhtrue could not be fit on page. It will not be plotted.

## Warning in wordcloud(rownames(article2), article2$article1, min.freq = 9, :
## classcsse9w26l could not be fit on page. It will not be plotted.

## Warning in wordcloud(rownames(article2), article2$article1, min.freq = 9, :
## classcssist4u3a could not be fit on page. It will not be plotted.
```


ariahiddentrue
propertyarticletag
news
classcssjq1cx6
classcss1jmp9xk
ariaexpandedfalse
331
reserve...
transparent
regulatory
classcssat9mc1
1857
review
bank
financial
said
changes
599px
powell
regulation
supervision
typebutton
jerome
shortcomings
swift
1px
edvi3so2a
later
classcssyywogo
viewbox0
101
mention
surrounding
bank...s
var
federal
events
...
thorough
demand
fill000
joint
smartphonea
lead
auto
statement
fed
new
york
function
printcss1gkjb1c
helped
solid
datatestidfooterlink
chair
silicon
true
reserve
height18
officials
valley
include
government
blocked
wanted
script
classcssa7htku
flaws
banking