

Week 9

Code-along

Slide 8

```
library(tidyverse)
suppressMessages(library(tidyverse))
```

```
tidydata <- tribble(
  ~country, ~year, ~cases, ~population,
  "Afghanistan", 1999, 745, 19987071,
  "Afghanistan", 2000, 2666, 20595360,
  "Brazil", 1999, 37737, 172006362,
  "Brazil", 2000, 80488, 174504898,
  "China", 1999, 212258, 1272915272,
  "China", 2000, 213766, 1280428583)
```

tidydata

```
## # A tibble: 6 x 4
##   country      year  cases population
##   <chr>      <dbl> <dbl>      <dbl>
## 1 Afghanistan 1999     745  19987071
## 2 Afghanistan 2000    2666  20595360
## 3 Brazil      1999   37737  172006362
## 4 Brazil      2000   80488  174504898
## 5 China       1999  212258  1272915272
## 6 China       2000  213766  1280428583
```

```
nontidydata <- tribble(
  ~country, ~year, ~rate,
  "Afghanistan", 1999, "745/19987071",
  "Afghanistan", 2000, "2666/20595360",
  "Brazil", 1999, "37737/172006362",
  "Brazil", 2000, "80488/174504898",
  "China", 1999, "212258/1272915272",
  "China", 2000, "213766/1280428583")
```

nontidydata

```
## # A tibble: 6 x 3
##   country      year rate
##   <chr>      <dbl> <chr>
## 1 Afghanistan 1999 745/19987071
## 2 Afghanistan 2000 2666/20595360
```

```
## 3 Brazil      1999 37737/172006362
## 4 Brazil      2000 80488/174504898
## 5 China       1999 212258/1272915272
## 6 China       2000 213766/1280428583
```

Slide 11

```
tidieddata <- nontidydata %>% separate(rate, into = c("cases", "population"), sep = "/")
tidieddata
```

```
## # A tibble: 6 x 4
##   country      year cases population
##   <chr>        <dbl> <chr>   <chr>
## 1 Afghanistan 1999 745    19987071
## 2 Afghanistan 2000 2666   20595360
## 3 Brazil      1999 37737  172006362
## 4 Brazil      2000 80488  174504898
## 5 China       1999 212258 1272915272
## 6 China       2000 213766 1280428583
```

Slide 12

```
newtidieddata <- tidieddata %>%
pivot_longer(
cols = cases:population,
names_to = "measurement",
values_to = "value"
)
newtidieddata
```

```
## # A tibble: 12 x 4
##   country      year measurement value
##   <chr>        <dbl> <chr>      <chr>
## 1 Afghanistan 1999 cases      745
## 2 Afghanistan 1999 population 19987071
## 3 Afghanistan 2000 cases      2666
## 4 Afghanistan 2000 population 20595360
## 5 Brazil      1999 cases      37737
## 6 Brazil      1999 population 172006362
## 7 Brazil      2000 cases      80488
## 8 Brazil      2000 population 174504898
## 9 China       1999 cases      212258
## 10 China      1999 population 1272915272
## 11 China      2000 cases      213766
## 12 China      2000 population 1280428583
```

Slide 14

```
df <- tribble(
  ~id, ~bp1, ~bp2,
  "A", 100, 120,
  "B", 140, 115,
  "C", 120, 125
)
df
```

```
## # A tibble: 3 x 3
##   id      bp1  bp2
##   <chr> <dbl> <dbl>
## 1 A      100   120
## 2 B      140   115
## 3 C      120   125
```

```
df %>%
  pivot_longer(
    cols = bp1:bp2,
    names_to = "measurement",
    values_to = "value"
  )
```

```
## # A tibble: 6 x 3
##   id      measurement value
##   <chr> <chr>         <dbl>
## 1 A      bp1             100
## 2 A      bp2             120
## 3 B      bp1             140
## 4 B      bp2             115
## 5 C      bp1             120
## 6 C      bp2             125
```

Slide 18

```
newtidieddata
```

```
## # A tibble: 12 x 4
##   country      year measurement value
##   <chr>      <dbl> <chr>      <chr>
## 1 Afghanistan 1999 cases      745
## 2 Afghanistan 1999 population 19987071
## 3 Afghanistan 2000 cases      2666
## 4 Afghanistan 2000 population 20595360
## 5 Brazil      1999 cases      37737
## 6 Brazil      1999 population 172006362
## 7 Brazil      2000 cases      80488
## 8 Brazil      2000 population 174504898
## 9 China       1999 cases      212258
## 10 China      1999 population 1272915272
## 11 China      2000 cases      213766
## 12 China      2000 population 1280428583
```

```
newtidieddata %>%
pivot_wider(names_from="measurement",
values_from="value")
```

```
## # A tibble: 6 x 4
##   country      year cases population
##   <chr>      <dbl> <chr>   <chr>
## 1 Afghanistan 1999  745   19987071
## 2 Afghanistan 2000 2666   20595360
## 3 Brazil      1999 37737  172006362
## 4 Brazil      2000 80488  174504898
## 5 China       1999 212258 1272915272
## 6 China       2000 213766 1280428583
```

Slide 19

```
df <- tribble(
~id, ~measurement, ~value,
"A", "bp1", 100,
"B", "bp1", 140,
"B", "bp2", 115,
"A", "bp2", 120,
"A", "bp3", 105
)
df
```

```
## # A tibble: 5 x 3
##   id      measurement value
##   <chr>   <chr>      <dbl>
## 1 A      bp1          100
## 2 B      bp1          140
## 3 B      bp2          115
## 4 A      bp2          120
## 5 A      bp3          105
```

```
df %>% pivot_wider(
  names_from = measurement,
  values_from = value
)
```

```
## # A tibble: 2 x 4
##   id      bp1      bp2      bp3
##   <chr> <dbl> <dbl> <dbl>
## 1 A      100    120    105
## 2 B      140    115     NA
```

Install and load rvest

```
library(rvest)
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:readr':
##
##   guess_encoding
```

```
webpage <- read_html("https://books.toscrape.com/")
table <- html_elements(webpage, "body")
```

```
library(httr)
library(jsonlite)
```

```
##
## Attaching package: 'jsonlite'
```

```
## The following object is masked from 'package:purrr':
##
##   flatten
```

```
# current data
current_county_data_url <- "https://api.covidactnow.org/v2/counties.csv?apiKey=YOUR_KEY_HERE"
# historic data
historic_county_data_url <- "https://api.covidactnow.org/v2/counties.timeseries.csv?apiKey=YOUR_KEY_HERE"
# individual location data
individual_loc_data_url <- "https://api.covidactnow.org/v2/county/{state}.csv?apiKey=YOUR_KEY_HERE"
```

Calling an API:

```
current_county_data_url <- "https://api.covidactnow.org/v2/counties.csv?apiKey=33382de96fd8441fb6c"
raw_data <- GET(current_county_data_url)
raw_data$status
```

```
## [1] 403
```

```
raw_data$content
```

```
## [1] 7b 22 65 72 72 6f 72 22 3a 20 22 49 6e 76 61 6c 69 64 20 41 50 49 20 6b 65
## [26] 79 2e 22 7d
```

```
individual_loc_data_url <-
"https://api.covidactnow.org/v2/county/{49}.csv?apiKey=33382de96fd8441fb6c1eca82b3bd4ec"
raw_data2 <- GET(individual_loc_data_url)
raw_data2$status
```

```
## [1] 403
```

```
raw_data2$content
```

```
## [1] 3c 3f 78 6d 6c 20 76 65 72 73 69 6f 6e 3d 22 31 2e 30 22 20 65 6e 63 6f 64
## [26] 69 6e 67 3d 22 55 54 46 2d 38 22 3f 3e 0a 3c 45 72 72 6f 72 3e 3c 43 6f 64
## [51] 65 3e 41 63 63 65 73 73 44 65 6e 69 65 64 3c 2f 43 6f 64 65 3e 3c 4d 65 73
```

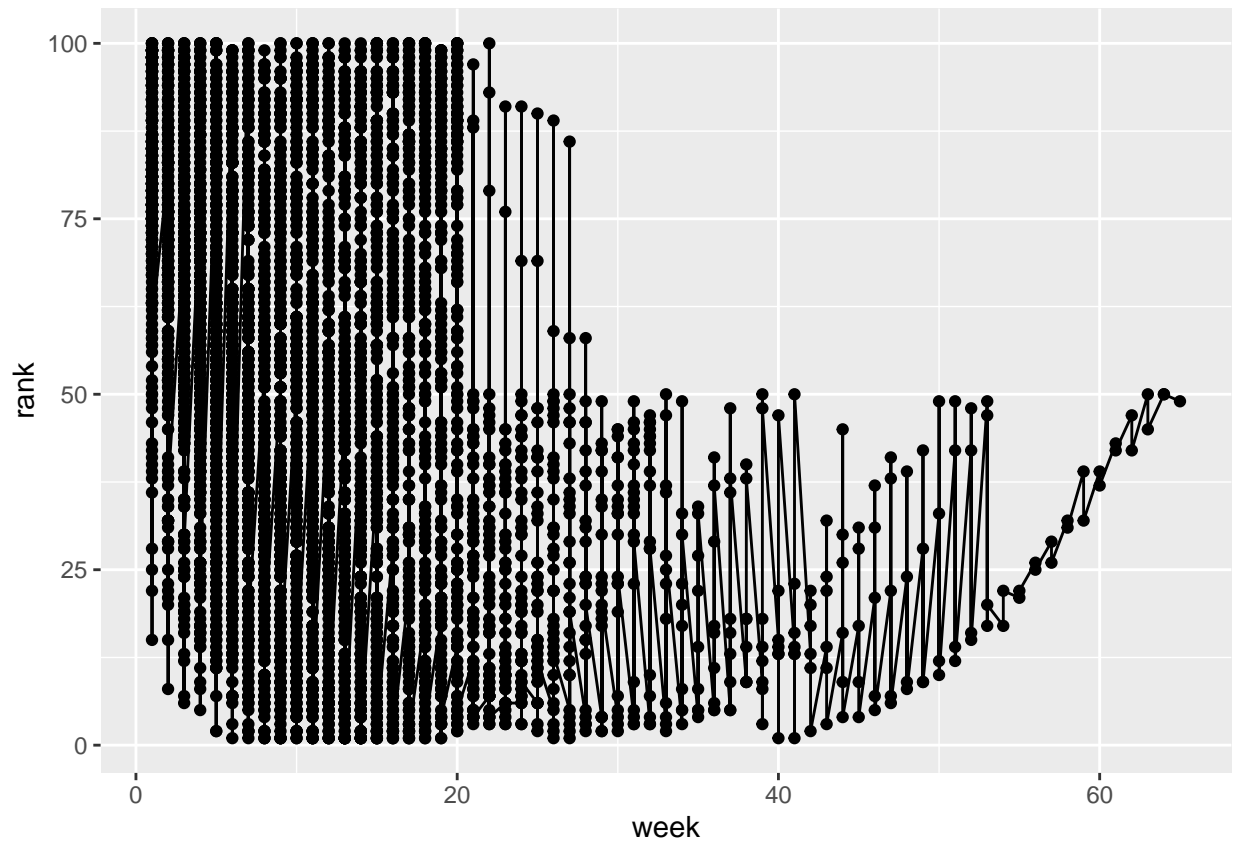
```
## [76] 73 61 67 65 3e 41 63 63 65 73 73 20 44 65 6e 69 65 64 3c 2f 4d 65 73 73 61
## [101] 67 65 3e 3c 52 65 71 75 65 73 74 49 64 3e 41 46 52 36 44 54 44 33 4d 34 54
## [126] 4b 41 58 47 4e 3c 2f 52 65 71 75 65 73 74 49 64 3e 3c 48 6f 73 74 49 64 3e
## [151] 6a 50 42 45 58 66 33 57 59 6e 69 65 51 57 48 6e 71 2f 59 7a 34 2f 34 33 31
## [176] 57 4b 44 38 2b 59 4a 5a 4d 6e 47 4d 62 77 4c 79 59 6c 2f 34 69 51 33 73 75
## [201] 2f 56 74 69 6b 59 69 34 69 6d 39 50 55 6a 57 6f 53 65 56 43 61 31 66 73 63
## [226] 3d 3c 2f 48 6f 73 74 49 64 3e 3c 2f 45 72 72 6f 72 3e
```

Challenge 1

```
pivot_longer_billboard <- billboard %>%
pivot_longer(
  cols = starts_with("wk"),
  names_to = "week",
  values_to = "rank",
  values_drop_na = TRUE
)

pivot_longer_billboard <- pivot_longer_billboard %>% mutate(week = parse_number(week))

ggplot(pivot_longer_billboard) +
  aes(x = week,
      y = rank) +
  geom_point() +
  geom_line()
```



Challenge 2

```
pivot_wider_cms <- cms_patient_experience %>%
pivot_wider(names_from="measure_cd",
values_from="prf_rate")
```

```
pivot_wider_cms <- cms_patient_experience %>%
pivot_wider(names_from="measure_cd",
values_from="prf_rate",
id_cols = starts_with("org"))
```

```
pivot_wider_cms
```

```
## # A tibble: 95 x 8
##   org_pac_id org_nm CAHPS_GRP_1 CAHPS_GRP_2 CAHPS_GRP_3 CAHPS_GRP_5 CAHPS_GRP_8
##   <chr>      <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 0446157747 USC C~         63         87         86         57         85
## 2 0446162697 ASSOC~         59         85         83         63         88
## 3 0547164295 BEAVE~         49         NA         75         44         73
## 4 0749333730 CAPE ~         67         84         85         65         82
## 5 0840104360 ALLIA~         66         87         87         64         87
```

##	6	0840109864	REX H~	73	87	84	67	91
##	7	0840513552	SCL H~	58	83	76	58	78
##	8	0941545784	GRITM~	46	86	81	54	NA
##	9	1052612785	COMMU~	65	84	80	58	87
##	10	1254237779	OUR L~	61	NA	NA	65	NA

i 85 more rows

i 1 more variable: CAHPS_GRP_12 <dbl>