Week-9: Code-along and challenge

NM2207: Computational Media Literacy

2023-10-16

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## 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
```

Code-along

```r
# Load data set
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)
```

```r
# Org by total # of cases in a year
tidydata %>%
 group_by(year) %>%
 summarize(total_cases = sum(cases))
```

```
## # A tibble: 2 x 2
##    year total_cases
##   <dbl>       <dbl>
## 1  1999      250740
## 2  2000      296920
```

```r
# Loading a non-tidy dataset
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
```

```r
# tidying the dataset
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
```

```r
# Passing names to new columns "measurement" and "values" from col "cases" to "popn"
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
```

```r
# one-dimensional ggplot
ggplot(newtidieddata) +
 aes(x=year,y=value, colour=country) +
 geom_point() +
 geom_line(aes(group = country))+
 facet_wrap(~measurement) +
 theme_bw()
```
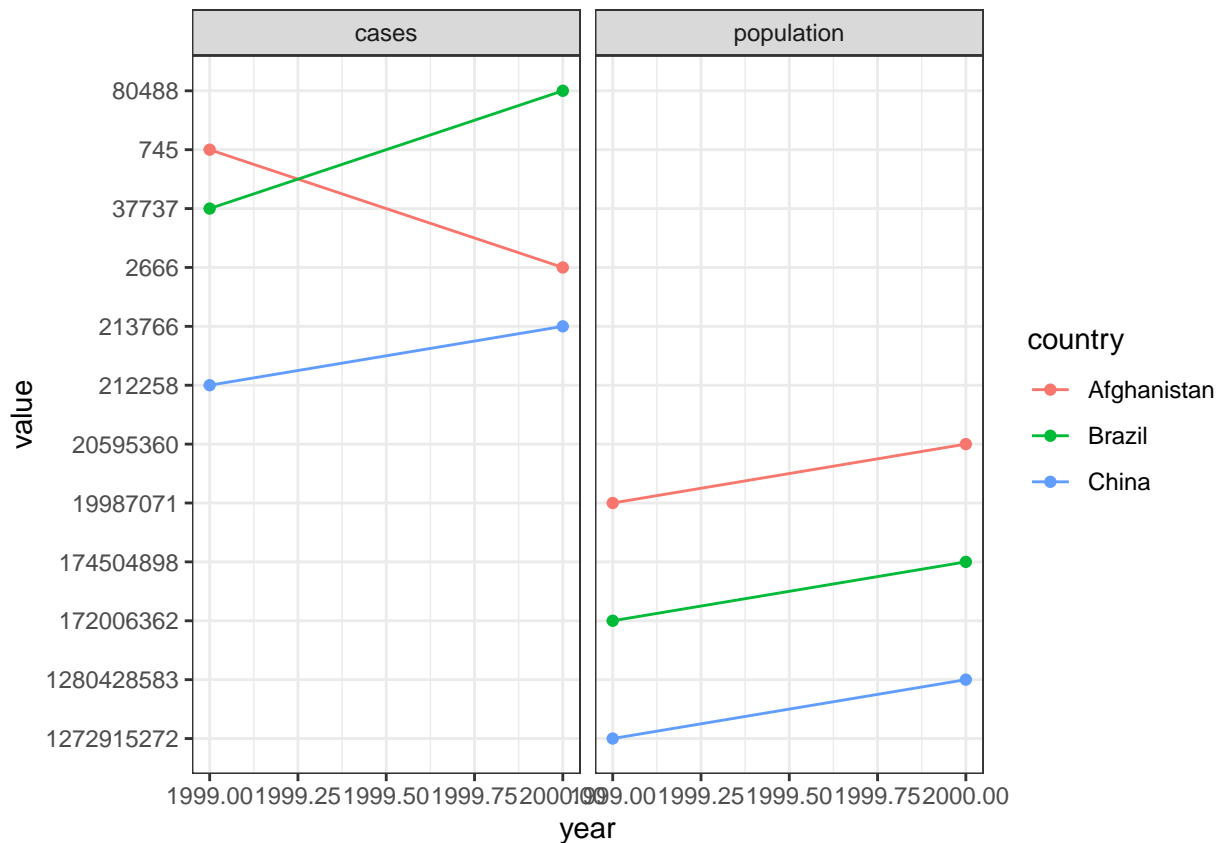


```r
# Example 2
df <- tribble(
 ~id, ~bp1, ~bp2,
 "A", 100, 120,
```

```
  "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
```

```
# from columns bp1 to bp2, send them to "measurement" and "value"
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
```

```
# going back to pivot_longer
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
```

```
# Example 4: reshapping data
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
```

```r
# reshapping data w pivot_wider
# NA appears when there is a missing entry
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
```

Challenge

Question 1

```r
# Load tidyverse package
library(tidyverse)
```

```r
# Pivot longer to arrange the names, values, and clean data
newtidieweek <- billboard %>%
  pivot_longer(
    cols = starts_with("wk"),
    names_to = "week",
    values_to = "rank",
    values_drop_na = TRUE,
  )

newtidieweek
```

```
## # A tibble: 5,307 x 5
##    artist track                date.entered week   rank
##    <chr>  <chr>                <date>       <chr> <dbl>
## 1 2 Pac   Baby Don't Cry (Keep... 2000-02-26   wk1     87
## 2 2 Pac   Baby Don't Cry (Keep... 2000-02-26   wk2     82
## 3 2 Pac   Baby Don't Cry (Keep... 2000-02-26   wk3     72
## 4 2 Pac   Baby Don't Cry (Keep... 2000-02-26   wk4     77
## 5 2 Pac   Baby Don't Cry (Keep... 2000-02-26   wk5     87
## 6 2 Pac   Baby Don't Cry (Keep... 2000-02-26   wk6     94
## 7 2 Pac   Baby Don't Cry (Keep... 2000-02-26   wk7     99
```

```
##  8 2Ge+her The Hardest Part Of ... 2000-09-02   wk1      91
##  9 2Ge+her The Hardest Part Of ... 2000-09-02   wk2      87
## 10 2Ge+her The Hardest Part Of ... 2000-09-02   wk3      92
## # i 5,297 more rows
```

```r
# mutate data set
tidie_data <- newtidieweek %>%
  mutate(week = parse_number(week))

tidie_data
```

```
## # A tibble: 5,307 x 5
##    artist  track                 date.entered  week  rank
##    <chr>   <chr>                 <date>       <dbl> <dbl>
##  1 2 Pac   Baby Don't Cry (Keep... 2000-02-26     1    87
##  2 2 Pac   Baby Don't Cry (Keep... 2000-02-26     2    82
##  3 2 Pac   Baby Don't Cry (Keep... 2000-02-26     3    72
##  4 2 Pac   Baby Don't Cry (Keep... 2000-02-26     4    77
##  5 2 Pac   Baby Don't Cry (Keep... 2000-02-26     5    87
##  6 2 Pac   Baby Don't Cry (Keep... 2000-02-26     6    94
##  7 2 Pac   Baby Don't Cry (Keep... 2000-02-26     7    99
##  8 2Ge+her The Hardest Part Of ... 2000-09-02     1    91
##  9 2Ge+her The Hardest Part Of ... 2000-09-02     2    87
## 10 2Ge+her The Hardest Part Of ... 2000-09-02     3    92
## # i 5,297 more rows
```

```r
# Plot graph
ggplot(tidie_data) +
 aes(x=week,y=rank) +
  geom_line()
```
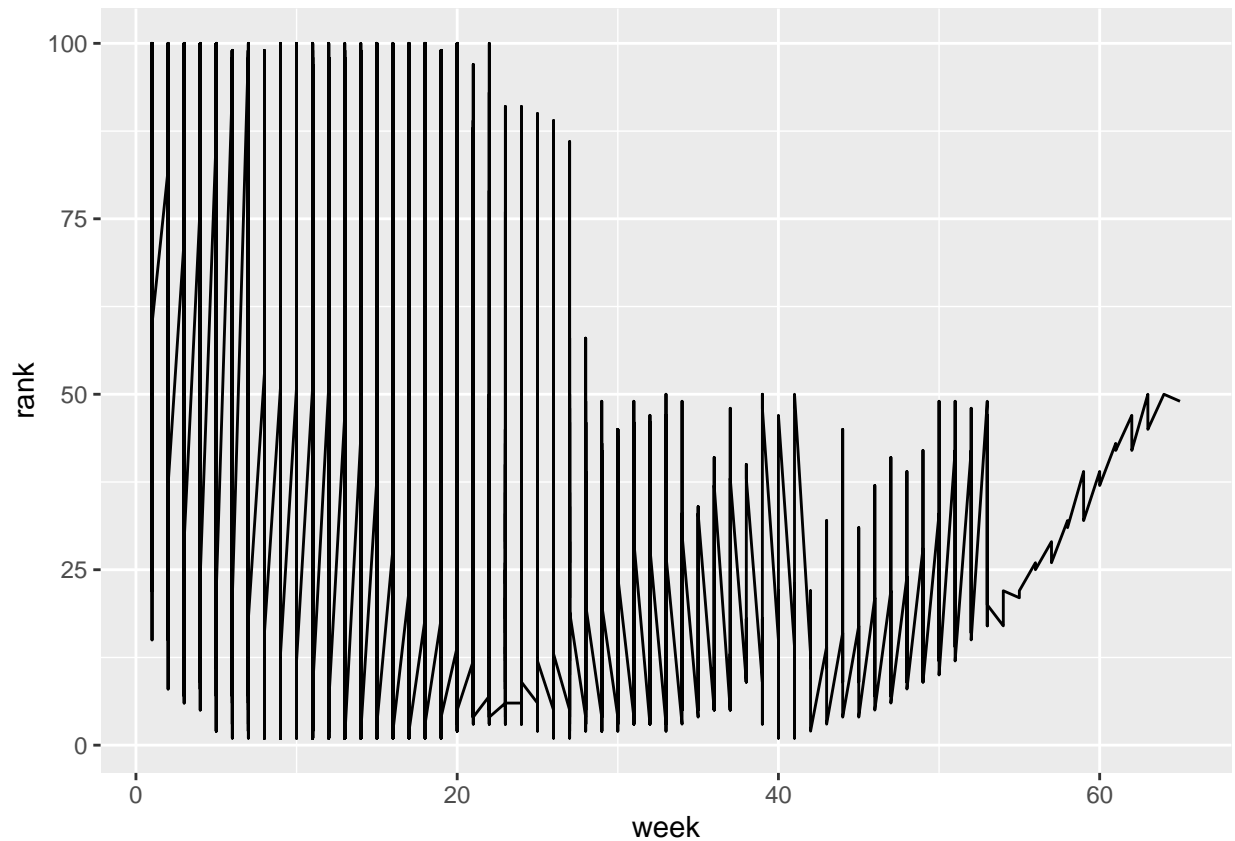
Question 2

```r
# use pivot_wider
new_measure_cd <- cms_patient_experience %>%
  pivot_wider(names_from = "measure_cd",
              values_from = "prf_rate",
              id_cols = starts_with("org"))

new_measure_cd
```

```
## # 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>
```