Challenge-9

Marzuki Nooranas

2023-10-16

## Code Along - 9

## Slide 8:

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.3 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

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 × 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 × 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:

nontidydata

## # A tibble: 6 × 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

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

## # A tibble: 6 × 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

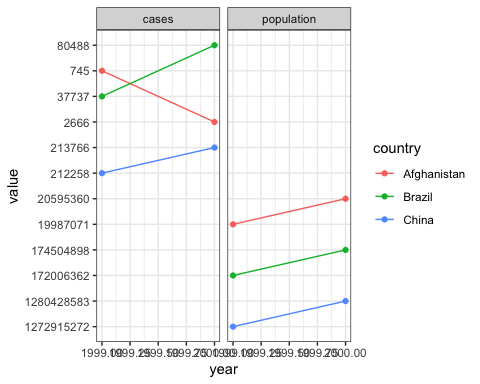
#Slide 12:

newtidieddata <- tidieddata %>%  
 pivot\_longer(  
 cols = cases:population,  
 names\_to = "measurement",  
 values\_to = "value"  
 )  
newtidieddata

## # A tibble: 12 × 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 13:

ggplot(newtidieddata) +  
 aes(x=year,y=value, colour=country) +  
 geom\_point() +  
 geom\_line(aes(group = country))+  
 facet\_wrap(~measurement) +  
 theme\_bw()



#Slide 14:

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

## # A tibble: 3 × 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 × 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 × 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 × 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 × 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  
)

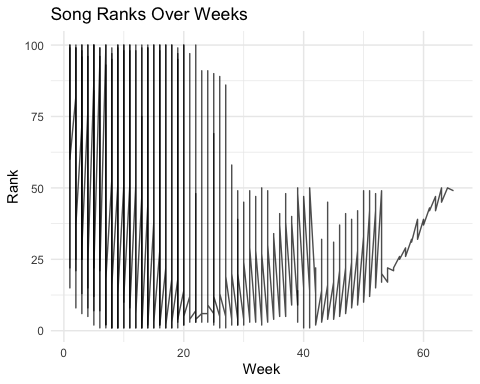
## # A tibble: 2 × 4  
## id bp1 bp2 bp3  
## <chr> <dbl> <dbl> <dbl>  
## 1 A 100 120 105  
## 2 B 140 115 NA

## Challenge 9 Question 1:

billboard

## # A tibble: 317 × 79  
## artist track date.entered wk1 wk2 wk3 wk4 wk5 wk6 wk7 wk8  
## <chr> <chr> <date> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2 Pac Baby… 2000-02-26 87 82 72 77 87 94 99 NA  
## 2 2Ge+her The … 2000-09-02 91 87 92 NA NA NA NA NA  
## 3 3 Doors D… Kryp… 2000-04-08 81 70 68 67 66 57 54 53  
## 4 3 Doors D… Loser 2000-10-21 76 76 72 69 67 65 55 59  
## 5 504 Boyz Wobb… 2000-04-15 57 34 25 17 17 31 36 49  
## 6 98^0 Give… 2000-08-19 51 39 34 26 26 19 2 2  
## 7 A\*Teens Danc… 2000-07-08 97 97 96 95 100 NA NA NA  
## 8 Aaliyah I Do… 2000-01-29 84 62 51 41 38 35 35 38  
## 9 Aaliyah Try … 2000-03-18 59 53 38 28 21 18 16 14  
## 10 Adams, Yo… Open… 2000-08-26 76 76 74 69 68 67 61 58  
## # ℹ 307 more rows  
## # ℹ 68 more variables: wk9 <dbl>, wk10 <dbl>, wk11 <dbl>, wk12 <dbl>,  
## # wk13 <dbl>, wk14 <dbl>, wk15 <dbl>, wk16 <dbl>, wk17 <dbl>, wk18 <dbl>,  
## # wk19 <dbl>, wk20 <dbl>, wk21 <dbl>, wk22 <dbl>, wk23 <dbl>, wk24 <dbl>,  
## # wk25 <dbl>, wk26 <dbl>, wk27 <dbl>, wk28 <dbl>, wk29 <dbl>, wk30 <dbl>,  
## # wk31 <dbl>, wk32 <dbl>, wk33 <dbl>, wk34 <dbl>, wk35 <dbl>, wk36 <dbl>,  
## # wk37 <dbl>, wk38 <dbl>, wk39 <dbl>, wk40 <dbl>, wk41 <dbl>, wk42 <dbl>, …

billboard\_new <- billboard %>%  
 pivot\_longer(cols = starts\_with("wk"),  
 names\_to = "week",  
 values\_to = "rank",  
 values\_drop\_na = TRUE) %>%  
 mutate(week = parse\_number(week))  
  
ggplot(billboard\_new, aes(x = week, y = rank)) +  
 geom\_line(alpha = 0.7) +   
 labs(title = "Song Ranks Over Weeks",   
 x = "Week",   
 y = "Rank") +  
 theme\_minimal()



## Challenge 9 Question 2:

cms\_patient\_experience

## # A tibble: 500 × 5  
## org\_pac\_id org\_nm measure\_cd measure\_title prf\_rate  
## <chr> <chr> <chr> <chr> <dbl>  
## 1 0446157747 USC CARE MEDICAL GROUP INC CAHPS\_GRP… CAHPS for MI… 63  
## 2 0446157747 USC CARE MEDICAL GROUP INC CAHPS\_GRP… CAHPS for MI… 87  
## 3 0446157747 USC CARE MEDICAL GROUP INC CAHPS\_GRP… CAHPS for MI… 86  
## 4 0446157747 USC CARE MEDICAL GROUP INC CAHPS\_GRP… CAHPS for MI… 57  
## 5 0446157747 USC CARE MEDICAL GROUP INC CAHPS\_GRP… CAHPS for MI… 85  
## 6 0446157747 USC CARE MEDICAL GROUP INC CAHPS\_GRP… CAHPS for MI… 24  
## 7 0446162697 ASSOCIATION OF UNIVERSITY PHYSI… CAHPS\_GRP… CAHPS for MI… 59  
## 8 0446162697 ASSOCIATION OF UNIVERSITY PHYSI… CAHPS\_GRP… CAHPS for MI… 85  
## 9 0446162697 ASSOCIATION OF UNIVERSITY PHYSI… CAHPS\_GRP… CAHPS for MI… 83  
## 10 0446162697 ASSOCIATION OF UNIVERSITY PHYSI… CAHPS\_GRP… CAHPS for MI… 63  
## # ℹ 490 more rows

cms\_wider <- cms\_patient\_experience %>%  
 pivot\_wider(names\_from = measure\_cd,  
 values\_from = prf\_rate,  
 id\_cols = starts\_with("org"))  
  
head(cms\_wider)

## # A tibble: 6 × 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 CA… 63 87 86 57 85  
## 2 0446162697 ASSOCI… 59 85 83 63 88  
## 3 0547164295 BEAVER… 49 NA 75 44 73  
## 4 0749333730 CAPE P… 67 84 85 65 82  
## 5 0840104360 ALLIAN… 66 87 87 64 87  
## 6 0840109864 REX HO… 73 87 84 67 91  
## # ℹ 1 more variable: CAHPS\_GRP\_12 <dbl>