

Week 9 Code along and Challenge

Chloe Tan

2023-10-16

R Markdown

```
#library(dplyr)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.3
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
## Warning: package 'readr' was built under R version 4.2.3
```

```
## Warning: package 'purrr' was built under R version 4.2.3
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
## Warning: package 'stringr' was built under R version 4.2.2
```

```
## Warning: package 'forcats' was built under R version 4.2.3
```

```
## Warning: package 'lubridate' was built under R version 4.2.3
```

```
## — 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 be
come errors
```

#slide 8

```
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

#slide 8

```
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

#slide 8 untidy data

```
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
```

```
#tidy non tidy data
#slide 11
#library(tidyr)

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

tidiedata
```

```
## # 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

newtidiedata<-tidiedata%>%pivot_longer(cols=cases:population,names_to="measurement",values_to
= "value")

newtidiedata
```



```
## # 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
```

#make cases and population column renames to new column called value while the cases and population are split into 2 categories under the names column so we get 4 country names instead of 2

#reshape data to suit our purpose, summarise cases in different years

```
#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
```

```
#piping slide 14
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
```

so we organise data to see which data from A is bp1 and bp2 as compared to the first example where bp1 and bp2 are combined together. so more specific data is produced.

```
#slide 18
newtidiedata
```

```
## # 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 18
newtidiedata%>%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
```

#go back to the original plot using pivot_wider

```
#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
```

```
#slide 19
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
```

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

```
rank<-billboard%>%pivot_longer(cols=wk1:wk76,names_to = "week",values_to = "value",values_drop_na = TRUE) %>% mutate(week=parse_number(week))
rank
```

```
## # A tibble: 5,307 × 5
##   artist      track      date.entered  week value
##   <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
```

```
ggplot(rank)+aes(x="artist",y="value")+geom_point()+geom_line(aes(group=artist))+facet_wrap(~value)+theme_bw()
```

	1	2	3	4	5	6	7	8	9	10
value	•	•	•	•	•	•	•	•	•	•
	11	12	13	14	15	16	17	18	19	20
value	•	•	•	•	•	•	•	•	•	•
	21	22	23	24	25	26	27	28	29	30
value	•	•	•	•	•	•	•	•	•	•
	31	32	33	34	35	36	37	38	39	40
value	•	•	•	•	•	•	•	•	•	•
	41	42	43	44	45	46	47	48	49	50
value	•	•	•	•	•	•	•	•	•	•
y	51	52	53	54	55	56	57	58	59	60
value	•	•	•	•	•	•	•	•	•	•
	61	62	63	64	65	66	67	68	69	70
value	•	•	•	•	•	•	•	•	•	•
	71	72	73	74	75	76	77	78	79	80
value	•	•	•	•	•	•	•	•	•	•
	81	82	83	84	85	86	87	88	89	90
value	•	•	•	•	•	•	•	•	•	•
	91	92	93	94	95	96	97	98	99	100
value	•	•	•	•	•	•	•	•	•	•
	artist	artist	artist	artist	artist	artist	artist	artist	artist	artist
	x									

#challenge 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
```

```
patients<-cms_patient_experience%>%pivot_wider(names_from = "measure_cd",values_from = "org_p
ac_id")

patients
```



```
## # A tibble: 500 × 9
##   org_nm measure_title prf_rate CAHPS_GRP_1 CAHPS_GRP_2 CAHPS_GRP_3 CAHPS_GRP_5
##   <chr>  <chr>          <dbl> <chr>      <chr>      <chr>      <chr>
## 1 USC C... CAHPS for MI...      63 0446157747 <NA>      <NA>      <NA>
## 2 USC C... CAHPS for MI...      87 <NA>      0446157747 <NA>      <NA>
## 3 USC C... CAHPS for MI...      86 <NA>      <NA>      0446157747 <NA>
## 4 USC C... CAHPS for MI...      57 <NA>      <NA>      <NA>      0446157747
## 5 USC C... CAHPS for MI...      85 <NA>      <NA>      <NA>      <NA>
## 6 USC C... CAHPS for MI...      24 <NA>      <NA>      <NA>      <NA>
## 7 ASSOC... CAHPS for MI...      59 0446162697 <NA>      <NA>      <NA>
## 8 ASSOC... CAHPS for MI...      85 <NA>      0446162697 <NA>      <NA>
## 9 ASSOC... CAHPS for MI...      83 <NA>      <NA>      0446162697 <NA>
## 10 ASSOC... CAHPS for MI...      63 <NA>      <NA>      <NA>      0446162697
## # i 490 more rows
## # i 2 more variables: CAHPS_GRP_8 <chr>, CAHPS_GRP_12 <chr>
```

```
patients<-cms_patient_experience%>%pivot_wider(names_from = "measure_cd",values_from = "org_p
ac_id")%>%select(id_cols=starts_with("org"))
```

```
patients
```

```
## # A tibble: 500 × 1
##   id_cols
##   <chr>
## 1 USC CARE MEDICAL GROUP INC
## 2 USC CARE MEDICAL GROUP INC
## 3 USC CARE MEDICAL GROUP INC
## 4 USC CARE MEDICAL GROUP INC
## 5 USC CARE MEDICAL GROUP INC
## 6 USC CARE MEDICAL GROUP INC
## 7 ASSOCIATION OF UNIVERSITY PHYSICIANS
## 8 ASSOCIATION OF UNIVERSITY PHYSICIANS
## 9 ASSOCIATION OF UNIVERSITY PHYSICIANS
## 10 ASSOCIATION OF UNIVERSITY PHYSICIANS
## # i 490 more rows
```