p8105_hw5_ruijipan

ruijipan

2022-11-13

R Markdown

```
> library(tidyverse)
> filenames = list.files('./data/longitudinal-study')
> filepaths = paste('./data/longitudinal-study/',filenames,sep="")
> df = map_dfr(filepaths, read.csv, .id = "input")
> df
   input week_1 week_2 week_3 week_4 week_5 week_6 week_7 week_8
1      1   0.20  -1.31   0.66   1.96   0.23   1.09   0.05   1.94
2      2   1.13  -0.88   1.07   0.17  -0.83  -0.31   1.58   0.44
3      3   1.77   3.11   2.22   3.26   3.31   0.89   1.88   1.01
4      4   1.04   3.66   1.22   2.33   1.47   2.70   1.87   1.66
5      5   0.47  -0.58  -0.09  -1.37  -0.32  -2.17   0.45   0.48
6      6   2.37   2.50   1.59  -0.16   2.08   3.07   0.78   2.35
7      7   0.03   1.21   1.13   0.64   0.49  -0.12  -0.07   0.46
8      8  -0.08   1.42   0.09   0.36   1.18  -1.16   0.33  -0.44
9      9   0.08   1.24   1.44   0.41   0.95   2.75   0.30   0.03
10    10   2.14   1.15   2.52   3.44   4.26   0.97   2.73  -0.53
11    11   3.05   3.67   4.84   5.80   6.33   5.46   6.38   5.91
12    12  -0.84   2.63   1.64   2.58   1.24   2.32   3.11   3.78
13    13   2.15   2.08   1.82   2.84   3.36   3.61   3.37   3.74
14    14  -0.62   2.54   3.78   2.73   4.49   5.82   6.00   6.49
15    15   0.70   3.33   5.34   5.57   6.90   6.66   6.24   6.95
16    16   3.73   4.08   5.40   6.41   4.87   6.09   7.66   5.83
17    17   1.18   2.35   1.23   1.17   2.02   1.61   3.13   4.88
18    18   1.37   1.43   1.84   3.60   3.80   4.72   4.68   5.70
19    19  -0.40   1.08   2.66   2.70   2.80   2.64   3.51   3.27
20    20   1.09   2.80   2.80   4.30   2.25   6.57   6.09   4.64
```

```
> df = rename(df,c(arm=input))
> str_arm = strsplit(filenames,"\\.")
>
> var_arm = vector("list", length = 20)
> var_id = vector("list", length = 20)
> for(i in 1:20){
+   var_arm[[i]] = strsplit(str_arm[[i]][1],"_")[[1]][1]
+   var_id[[i]] = strsplit(str_arm[[i]][1],"_")[[1]][2]
+ }
> df$arm = var_arm
> df$subject_id = var_id
> df = df %>%
+   select(arm, subject_id, everything())
> df
```

```
   arm subject_id week_1 week_2 week_3 week_4 week_5 week_6 week_7 week_8
1  con         01   0.20  -1.31   0.66   1.96   0.23   1.09   0.05   1.94
2  con         02   1.13  -0.88   1.07   0.17  -0.83  -0.31   1.58   0.44
3  con         03   1.77   3.11   2.22   3.26   3.31   0.89   1.88   1.01
4  con         04   1.04   3.66   1.22   2.33   1.47   2.70   1.87   1.66
5  con         05   0.47  -0.58  -0.09  -1.37  -0.32  -2.17   0.45   0.48
6  con         06   2.37   2.50   1.59  -0.16   2.08   3.07   0.78   2.35
7  con         07   0.03   1.21   1.13   0.64   0.49  -0.12  -0.07   0.46
8  con         08  -0.08   1.42   0.09   0.36   1.18  -1.16   0.33  -0.44
9  con         09   0.08   1.24   1.44   0.41   0.95   2.75   0.30   0.03
10 con         10   2.14   1.15   2.52   3.44   4.26   0.97   2.73  -0.53
11 exp         01   3.05   3.67   4.84   5.80   6.33   5.46   6.38   5.91
12 exp         02  -0.84   2.63   1.64   2.58   1.24   2.32   3.11   3.78
13 exp         03   2.15   2.08   1.82   2.84   3.36   3.61   3.37   3.74
14 exp         04  -0.62   2.54   3.78   2.73   4.49   5.82   6.00   6.49
15 exp         05   0.70   3.33   5.34   5.57   6.90   6.66   6.24   6.95
16 exp         06   3.73   4.08   5.40   6.41   4.87   6.09   7.66   5.83
17 exp         07   1.18   2.35   1.23   1.17   2.02   1.61   3.13   4.88
18 exp         08   1.37   1.43   1.84   3.60   3.80   4.72   4.68   5.70
19 exp         09  -0.40   1.08   2.66   2.70   2.80   2.64   3.51   3.27
20 exp         10   1.09   2.80   2.80   4.30   2.25   6.57   6.09   4.64
```

```
> homicides_df = read.csv("./data/homicides/homicide-data.csv")
> homicides_df$cit_state = paste(homicides_df$city,homicides_df$state)
> head(homicides_df)
         uid reported_date victim_last victim_first victim_race victim_age
1 Alb-000001      20100504      GARCIA         JUAN    Hispanic         78
2 Alb-000002      20100216     MONTOYA      CAMERON    Hispanic         17
3 Alb-000003      20100601 SATTERFIELD      VIVIANA       White         15
4 Alb-000004      20100101    MENDIOLA       CARLOS    Hispanic         32
5 Alb-000005      20100102        MULA       VIVIAN       White         72
6 Alb-000006      20100126        BOOK    GERALDINE       White         91
  victim_sex        city state      lat          lon           disposition
1       Male Albuquerque    NM 35.09579 -106.5385549 Closed without arrest
2       Male Albuquerque    NM 35.05681  -106.715321      Closed by arrest
3     Female Albuquerque    NM 35.08609  -106.695568 Closed without arrest
4       Male Albuquerque    NM 35.07849 -106.5560938      Closed by arrest
5     Female Albuquerque    NM 35.13036 -106.5809862 Closed without arrest
6     Female Albuquerque    NM 35.15111  -106.537797        Open/No arrest
       cit_state
1 Albuquerque NM
2 Albuquerque NM
3 Albuquerque NM
4 Albuquerque NM
5 Albuquerque NM
6 Albuquerque NM
```

```
> homicidesBycity =
+ homicides_df %>%
+   group_by(city) %>%
+   summarise(count = n())
> head(homicidesBycity)
# A tibble: 6 x 2
  city        count
```

```
  <chr>       <int>
1 Albuquerque   378
2 Atlanta       973
3 Baltimore    2827
4 Baton Rouge   424
5 Birmingham    800
6 Boston        614
```