Tidyverse Homework 3

Jessie Heise

```r
# Load packages needed for assignment
library(tidyverse)
```

Warning: package 'tidyverse' was built under R version 4.4.3

Warning: package 'ggplot2' was built under R version 4.4.3

Warning: package 'tibble' was built under R version 4.4.3

Warning: package 'tidyr' was built under R version 4.4.3

Warning: package 'readr' was built under R version 4.4.3

Warning: package 'purrr' was built under R version 4.4.3

Warning: package 'dplyr' was built under R version 4.4.3

Warning: package 'stringr' was built under R version 4.4.3

Warning: package 'forcats' was built under R version 4.4.3

Warning: package 'lubridate' was built under R version 4.4.3

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.2     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.4
-- 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 becom
```

```r
library(palmerpenguins)
```

```
Warning: package 'palmerpenguins' was built under R version 4.4.3
```

```r
## Task 1
### Part a
#### We cannot use specifically the read.csv() to read in these data because
#### the delimiter is a semicolon instead of a comma. Instead we must use
#### read.csv2()
# Read in data
read_csv2("Homework3data/data.txt", col_names = TRUE)
```

```
i Using "','" as decimal and "'.'" as grouping mark. Use `read_delim()` for more control.


Rows: 2 Columns: 3
-- Column specification -----------------------------------------------------
Delimiter: ";"
dbl (3): x, y, z

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.


# A tibble: 2 x 3
      x     y     z
  <dbl> <dbl> <dbl>
1     1     2     3
2     5     3     8
```

```
### Part b
# Read in data2
read_delim("Homework3data/data2.txt", delim = "6", col_types = "fdc")
```

```
# A tibble: 3 x 3
  x         y z
  <fct> <dbl> <chr>
1 1         2 3
2 5         3 8
3 7         4 2
```

```
## Task 2
### Part a
# Read in trailblazers data
trailblazer <- read_csv("Homework3data/trailblazer.csv")
```

```
Rows: 9 Columns: 11
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (1): Player
dbl (10): Game1_Home, Game2_Home, Game3_Away, Game4_Home, Game5_Home, Game6_...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Make sure data read in correctly
glimpse(trailblazer)
```

```
Rows: 9
Columns: 11
$ Player      <chr> "Damian Lillard", "CJ McCollum", "Norman Powell", "Robert ~
$ Game1_Home  <dbl> 20, 24, 14, 8, 20, 5, 11, 2, 7
$ Game2_Home  <dbl> 19, 28, 16, 6, 9, 5, 18, 8, 11
$ Game3_Away  <dbl> 12, 20, NA, 0, 4, 8, 12, 5, 5
$ Game4_Home  <dbl> 20, 25, NA, 3, 17, 10, 17, 8, 9
$ Game5_Home  <dbl> 25, 14, 12, 9, 14, 9, 5, 3, 8
$ Game6_Away  <dbl> 14, 25, 14, 6, 13, 6, 19, 8, 8
$ Game7_Away  <dbl> 20, 20, 22, 0, 7, 0, 17, 7, 4
$ Game8_Away  <dbl> 26, 21, 23, 6, 6, 7, 15, 0, 0
$ Game9_Home  <dbl> 4, 27, 25, 19, 10, 0, 16, 2, 7
$ Game10_Home <dbl> 25, 7, 13, 12, 15, 6, 10, 4, 8
```

```
### Part b
# Pivot & tidy the data
trailblazer_longer <-
  pivot_longer(
    trailblazer,
    cols = starts_with("Game"),
    names_to = "Game",
    values_to = "Points"
    )
trailblazer_longer <-
  separate(
    trailblazer_longer,
    Game,
    into = c("Game", "Location"),
    sep = "_"
    )

head(trailblazer_longer, n = 5)
```

```
# A tibble: 5 x 4
  Player          Game  Location Points
  <chr>           <chr> <chr>     <dbl>
1 Damian Lillard Game1 Home         20
2 Damian Lillard Game2 Home         19
3 Damian Lillard Game3 Away         12
4 Damian Lillard Game4 Home         20
5 Damian Lillard Game5 Home         25
```

```
### Part c
# Answer which players scored more, on average, when playing at home versus
# away in a single pipline
trailblazer_wider <-
  pivot_wider(trailblazer_longer,
              values_from = "Points",
              names_from = "Location") %>%
  group_by(Player) %>%
  summarise(mean_home = mean(Home, na.rm = TRUE),
        mean_away = mean(Away, na.rm = TRUE)) %>%
  mutate(mean_diff = mean_home - mean_away) %>%
  arrange(desc(mean_diff))
print(trailblazer_wider)
```

```
# A tibble: 9 x 4
  Player           mean_home mean_away mean_diff
  <chr>                <dbl>     <dbl>     <dbl>
1 Jusuf Nurkic          14.2       7.5     6.67
2 Robert Covington       9.5       3       6.5
3 Nassir Little          8.33      4.25    4.08
4 Damian Lillard        18.8      18       0.833
5 Cody Zeller            5.83      5.25    0.583
6 Larry Nance Jr         4.5       5      -0.5
7 CJ McCollum           20.8      21.5    -0.667
8 Anfernee Simons       12.8      15.8    -2.92
9 Norman Powell         16        19.7    -3.67
```

```
# Jusuf Nurkic, Robert Covington, Nassir Little, Damian Lillard,
# and Cody Zeller scored, on average, more points at home through
# the first 10 games of the season than away.
```

```
## Task 3
### Part a
#### <NULL> is saying that there are is no list for that species an island.
#### <dbl [52]> is describing that there is a vector of variables that are doubles.
#### <list> is describing that the output tibble contains a list.
### Part b
penguins %>%
  select(species, island, bill_length_mm) %>%
  count (species, island) %>%
  pivot_wider(values_from = n, names_from = island, values_fill = 0) %>%
  mutate(across(c(Biscoe, Dream, Torgersen), as.double))
```

```
# A tibble: 3 x 4
  species   Biscoe Dream Torgersen
  <fct>      <dbl> <dbl>     <dbl>
1 Adelie        44    56        52
2 Chinstrap      0    68         0
3 Gentoo       124     0         0
```

```
## Task 4
# Adelie is 26 and Gentoo is 30
penguins %>%
  mutate(
    bill_length_mm = ifelse(
```

```
      is.na(bill_length_mm),
      case_when(
        species == "Adelie" ~ 26,
        species == "Gentoo" ~ 30
        ),
      bill_length_mm
    )) %>%
  arrange(bill_length_mm) %>%
  head(penguins, n = 10)
```

```
# A tibble: 10 x 8
   species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
   <fct>   <fct>              <dbl>         <dbl>             <int>       <int>
 1 Adelie  Torgersen           26            NA                NA          NA
 2 Gentoo  Biscoe              30            NA                NA          NA
 3 Adelie  Dream             32.1          15.5               188        3050
 4 Adelie  Dream             33.1          16.1               178        2900
 5 Adelie  Torgersen         33.5          19                 190        3600
 6 Adelie  Dream             34            17.1               185        3400
 7 Adelie  Torgersen         34.1          18.1               193        3475
 8 Adelie  Torgersen         34.4          18.4               184        3325
 9 Adelie  Biscoe            34.5          18.1               187        2900
10 Adelie  Torgersen         34.6          21.1               198        4400
# i 2 more variables: sex <fct>, year <int>
```