ST558ProgrammingHW3

Holly Probasco

Task 1

part a

Here we don't use read_csv specifically because the delimiter is not a comma.

```
library(tidyverse)
```

```
-- 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 becor
```

```
data = read_csv2('data/data.txt')
```

```
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.
```

```
data
```

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

part b

```
data2 = read_delim('data/data2.txt', delim = "6", col_names = TRUE,
col_types = c("f","d","c"))
data2
```

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

Task 2

part a

```
trailblazer <- read_csv("data/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.
```

```
head(trailblazer)
```

```
# A tibble: 6 x 11
  Player       Game1_Home Game2_Home Game3_Away Game4_Home Game5_Home Game6_Away
  <chr>             <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1 Damian Lill~         20         19         12         20         25         14
2 CJ McCollum          24         28         20         25         14         25
3 Norman Powe~         14         16         NA         NA         12         14
4 Robert Covi~          8          6          0          3          9          6
5 Jusuf Nurkic         20          9          4         17         14         13
6 Cody Zeller           5          5          8         10          9          6
# i 4 more variables: Game7_Away <dbl>, Game8_Away <dbl>, Game9_Home <dbl>,
#   Game10_Home <dbl>
```

part b

```
trailblazer_longer = trailblazer |> pivot_longer(cols = 2:11,
names_to = c('Game', 'Location'), names_sep = "_",
values_to = 'Points')
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

```
trailblazer_wider = trailblazer_longer |>
pivot_wider(names_from = Location, values_from = Points) |>
  group_by(Player) |>
mutate(mean_home = mean(Home, na.rm=TRUE), mean_away = mean(Away, na.rm = TRUE),
pt_diff = mean_home-mean_away) |>
  arrange(desc(pt_diff))
```

The players that scored more at home than away are Jusuf Nurkic , Robert Covington, and Nassir Little

Task 3

part a

```
library(palmerpenguins)
```

- means that there is no data there
- <dbl [52]> means this element is a double list of size 52
- <list> means that this is a column of lists

part b

```
penguins |>  summarise(n = n(), .by = c(species, island)) |>
  pivot_wider(names_from = island, values_from = n, values_fill = 0)
```

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

Task 4

```
penguins |> mutate(bill_length_mm = case_when(is.na(bill_length_mm)
   & species == "Adelie" ~ 26, is.na(bill_length_mm) & species ==
  "Gentoo" ~ 30, TRUE ~ 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>
```