

Homework 3 Output

Mike Maccia

Loading libraries

```
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 become
```

```
library(palmerpenguins)
```

Attaching package: 'palmerpenguins'

The following objects are masked from 'package:datasets':

penguins, penguins_raw

Task 1

Question a: Why `read_csv` cannot be used to read the `data.txt` file

The function `read_csv()` can only be used to read in files that use the delimiters of commas or tabs. The function `read_csv2()` must be used in files with semicolons as the separator (commas can be used for decimal points).

```
data <- read_csv2('~/.ST558 Repo/Homework3/Data/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.

```
data
```

```
# A tibble: 2 x 3
```

```
      x      y      z
  <dbl> <dbl> <dbl>
1     1     2     3
2     5     3     8
```

Question b: Reading in 2nd file

In this file, “6” is the delimiter.

```
data_2 <- read_delim('~/.ST558 Repo/Homework3/Data/data2.txt',
                     delim = '6',
                     col_types= 'fdc')
```

```
data_2
```

```
# 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

Data tidying skills

Question a: Reading Data

Reading in the `trailblazer.csv` data

```
trailblazer <- read_csv('~/.ST558 Repo/Homework3/Data/trailblazer.csv',  
                        col_names= TRUE)
```

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.

```
trailblazer
```

A tibble: 9 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
7	Anfernee Si~	11	18	12	17	5	19
8	Larry Nance~	2	8	5	8	3	8
9	Nassir Litt~	7	11	5	9	8	8

i 4 more variables: Game7_Away <dbl>, Game8_Away <dbl>, Game9_Home <dbl>,

Game10_Home <dbl>

Question b: Pivoting the data longer

```

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

print(trailblazer_longer, n=5)

```

```

# A tibble: 90 x 4
  Player      Game Location Points
  <chr>      <chr> <chr>    <dbl>
1 Damian Lillard 1     Home      20
2 Damian Lillard 2     Home      19
3 Damian Lillard 3     Away      12
4 Damian Lillard 4     Home      20
5 Damian Lillard 5     Home      25
# i 85 more rows

```

Question c: Who scored more when playing at home versus away

```

trailblazer_wider <- trailblazer_longer |>
  mutate(Game = as.integer(Game)) |>
  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),
         diff_points = (mean_home - mean_away)) |>
  arrange(desc(diff_points))

print(trailblazer_wider, n = 25)

```

```

# A tibble: 90 x 7
# Groups:   Player [9]
  Player      Game Home Away mean_home mean_away diff_points
  <chr>      <int> <dbl> <dbl>    <dbl>    <dbl>    <dbl>
1 Jusuf Nurkic      1    20    NA     14.2     7.5     6.67

```

2	Jusuf Nurkic	2	9	NA	14.2	7.5	6.67
3	Jusuf Nurkic	3	NA	4	14.2	7.5	6.67
4	Jusuf Nurkic	4	17	NA	14.2	7.5	6.67
5	Jusuf Nurkic	5	14	NA	14.2	7.5	6.67
6	Jusuf Nurkic	6	NA	13	14.2	7.5	6.67
7	Jusuf Nurkic	7	NA	7	14.2	7.5	6.67
8	Jusuf Nurkic	8	NA	6	14.2	7.5	6.67
9	Jusuf Nurkic	9	10	NA	14.2	7.5	6.67
10	Jusuf Nurkic	10	15	NA	14.2	7.5	6.67
11	Robert Covington	1	8	NA	9.5	3	6.5
12	Robert Covington	2	6	NA	9.5	3	6.5
13	Robert Covington	3	NA	0	9.5	3	6.5
14	Robert Covington	4	3	NA	9.5	3	6.5
15	Robert Covington	5	9	NA	9.5	3	6.5
16	Robert Covington	6	NA	6	9.5	3	6.5
17	Robert Covington	7	NA	0	9.5	3	6.5
18	Robert Covington	8	NA	6	9.5	3	6.5
19	Robert Covington	9	19	NA	9.5	3	6.5
20	Robert Covington	10	12	NA	9.5	3	6.5
21	Nassir Little	1	7	NA	8.33	4.25	4.08
22	Nassir Little	2	11	NA	8.33	4.25	4.08
23	Nassir Little	3	NA	5	8.33	4.25	4.08
24	Nassir Little	4	9	NA	8.33	4.25	4.08
25	Nassir Little	5	8	NA	8.33	4.25	4.08

i 65 more rows

While they did not necessarily score the most points, Jusuf Nurkic (6.67) and Robert Covington (6.5) scored on average more points at home than away through the first 10 games of the season.