

# HW3 - Intro to Tidyverse

Kayla Kippes

## Task 1

### Part A

We cannot use the `read_csv()` to read in these data because the delimiter is a semicolon and not a comma.

```
library(tidyverse)
library(palmerpenguins)

data <- read_csv2("data/data.txt")
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 = ";", col_types = "fdc")
data2
```

```
# 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

```
trailblazer <- read_csv("data/trailblazer.csv")
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>
```

### Part B

```
library(tidyr)
#Put the data into long format and tidy up columns
trailblazer_longer <- trailblazer |>
  pivot_longer(cols = 2:11,
               names_to = c("Game", "Location"),
               names_sep = "_",
               values_to = "Points")
trailblazer_longer[1:5,]
```

```
# 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_longer |>
  pivot_wider(names_from = Location, values_from = Points, values_fill = 0) |>
  group_by(Player) |>
  summarise(Mean_Home = mean(Home, na.rm = TRUE),
            Mean_Away = mean(Away, na.rm = TRUE),
            Difference = Mean_Home - Mean_Away) |>
  arrange(desc(Difference))
```

```
# A tibble: 9 x 4
  Player      Mean_Home Mean_Away Difference
  <chr>         <dbl>     <dbl>     <dbl>
1 Jusuf Nurkic      8.5         3         5.5
2 Robert Covington  5.7         1.2         4.5
3 Damian Lillard    11.3        7.2         4.1
4 CJ McCollum      12.5        8.6         3.9
5 Nassir Little      5         1.7         3.3
6 Norman Powell     8.89        6.56         2.33
7 Anfernee Simons    7.7         6.3         1.4
8 Cody Zeller       3.5         2.1         1.4
9 Larry Nance Jr    2.7         2          0.7
```

Jusuf Nurkic, Robert Covington, Nassir Little, Damian Lillard, and Cody Zeller all scored more points at home than away in the first 10 games.

## Task 3

### Part A

means that there are no values for that species and island combination.

<dbl [52]> means that the list contains 52 values in it.

means that the output contains a list-column.

## Part B

```
penguins |>
  select(species, island, bill_length_mm) |>
  group_by(species, island) |>
  summarise(count = n()) |>
  pivot_wider(names_from = island, values_from = count, values_fill = 0)
```

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

## Task 4

```
penguins |>
  select(species, bill_length_mm) |>
  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) |>
  print(n = 10)
```

```
# A tibble: 344 x 2
  species bill_length_mm
  <fct>     <dbl>
1 Adelie      26
2 Gentoo      30
3 Adelie     32.1
4 Adelie     33.1
5 Adelie     33.5
6 Adelie      34
7 Adelie     34.1
8 Adelie     34.4
9 Adelie     34.5
```

```
10 Adelie          34.6
# i 334 more rows
```