Programming in tidyverse

```r
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   3.5.1     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 becon
```

```
library(palmerpenguins)
```

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

Task 1

Part a

We cannot use `read_csv()` for the `data.txt` file because this data is separated by `;`, while `read_csv()` can only read in data with values that are separated by `,`. For the `data.txt` file, we can use `read_csv2()` to read in data that is separated by `;`.

```
data <- read_csv2("data/data.txt", col_types = "d", skip = 0)
```

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

```
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_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")
```

```
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>
```

Part b

```
trailblazer_longer <- trailblazer |>
  pivot_longer(cols = !Player,
               names_to = "Games_Location",
               values_to = "Points") |>
  separate_wider_delim(cols = Games_Location,
                       delim = "_",
                       names = c("Game", "Location"),
                       cols_remove = TRUE)
trailblazer_longer %>% slice_head(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))
trailblazer_wider
```

```
# A tibble: 90 x 6
# Groups:   Player [9]
   Player         Game   Home  Away mean_home mean_away
   <chr>          <chr> <dbl> <dbl>     <dbl>     <dbl>
 1 Damian Lillard Game1    20    NA      18.8        18
 2 Damian Lillard Game2     19    NA      18.8        18
 3 Damian Lillard Game3     NA    12      18.8        18
 4 Damian Lillard Game4     20    NA      18.8        18
 5 Damian Lillard Game5     25    NA      18.8        18
 6 Damian Lillard Game6     NA    14      18.8        18
 7 Damian Lillard Game7     NA    20      18.8        18
 8 Damian Lillard Game8     NA    26      18.8        18
 9 Damian Lillard Game9      4    NA      18.8        18
10 Damian Lillard Game10    25    NA      18.8        18
# i 80 more rows
```

```
trailblazer_wider |> group_by(Player) |>
  group_by(Player) |>
  mutate(mean_home = mean(Home, na.rm = TRUE),
         mean_away = mean(Away, na.rm = TRUE)) |>
  mutate(home_away_diff = mean_home - mean_away) |>
  arrange(desc(home_away_diff))
```

```
# A tibble: 90 x 7
# Groups:   Player [9]
   Player       Game    Home  Away mean_home mean_away home_away_diff
   <chr>        <chr>  <dbl> <dbl>     <dbl>     <dbl>          <dbl>
 1 Jusuf Nurkic Game1     20    NA      14.2       7.5           6.67
 2 Jusuf Nurkic Game2      9    NA      14.2       7.5           6.67
 3 Jusuf Nurkic Game3     NA     4      14.2       7.5           6.67
 4 Jusuf Nurkic Game4     17    NA      14.2       7.5           6.67
 5 Jusuf Nurkic Game5     14    NA      14.2       7.5           6.67
 6 Jusuf Nurkic Game6     NA    13      14.2       7.5           6.67
 7 Jusuf Nurkic Game7     NA     7      14.2       7.5           6.67
 8 Jusuf Nurkic Game8     NA     6      14.2       7.5           6.67
 9 Jusuf Nurkic Game9     10    NA      14.2       7.5           6.67
10 Jusuf Nurkic Game10    15    NA      14.2       7.5           6.67
# i 80 more rows
```

On average, Jusuf Nurkic scored more points at home than away through the first 10 games of the season as shown with the highest `home_away_diff`.

Task 3

Part a

- <NULL> This is showing that there is not a value that exists for this variable.

- <dbl [52]> This is showing that this variable contain 52 different values.
- <list> This shows that the column contains lists as its values.

Part b

```
colleague_penguins <- penguins |> select(species, island, bill_length_mm) |>
  group_by(species, island) |>
  summarise(n = n(), .groups = "drop") |>
  pivot_wider(names_from = island, values_from = n) |>
  mutate(across(.cols = -(1:1), .fns = ~replace_na(., replace = 0)))
colleague_penguins
```

```
# A tibble: 3 x 4
  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_na_rm <- penguins |> select(species, bill_length_mm) |>
  mutate(bill_length_mm = replace(bill_length_mm,
                                  is.na(bill_length_mm) & species == "Adelie",
                                  26),
         bill_length_mm = replace(bill_length_mm,
                                  is.na(bill_length_mm) & species == "Gentoo",
                                  30)) |>
  arrange(bill_length_mm)
penguins_na_rm %>% slice_head(n = 10)
```

```
# A tibble: 10 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
```