Warmup: data wrangling with the Mets

Group members:

Instructions: Work with a neighbor on the following activity. I will collect the handout at the end of class, and it will be part of your class participation grade. You will be graded only on effort – it is ok if you don't finish all the questions, or get them all correct.

Baseball data

The `Teams` dataset in contains 3045 rows and 48 columns, and describes yearly statistics for baseball teams between 1871 and 2023.

Here are the first few rows and columns of the dataset:

```
  yearID lgID teamID franchID divID Rank  G Ghome  W  L DivWin WCWin LgWin
1   1871   NA    BS1      BNA  <NA>    3 31    NA 20 10   <NA>  <NA>     N
2   1871   NA    CH1      CNA  <NA>    2 28    NA 19  9   <NA>  <NA>     N
3   1871   NA    CL1      CFC  <NA>    8 29    NA 10 19   <NA>  <NA>     N
4   1871   NA    FW1      KEK  <NA>    7 19    NA  7 12   <NA>  <NA>     N
5   1871   NA    NY2      NNA  <NA>    5 33    NA 16 17   <NA>  <NA>     N
6   1871   NA    PH1      PNA  <NA>    1 28    NA 21  7   <NA>  <NA>     Y
```

Questions

1. The following chunk of code is run, producing the output below. Explain what the `select` function is doing.

```
Teams |>
  select(yearID, franchID, W, L)
```

```
  yearID franchID  W  L
1   1871      BNA 20 10
2   1871      CNA 19  9
3   1871      CFC 10 19
4   1871      KEK  7 12
5   1871      NNA 16 17
6   1871      PNA 21  7
```

2. Now suppose we want to get *only* the data for the New York Mets (franchise ID "NYM"), between 1998 and 2018. Fill in the code; the first few rows of the desired output are shown below.

```
Teams |>
  select(yearID, franchID, W, L) |>
  filter(...)
```

```
  yearID franchID  W  L
1   1998      NYM 88 74
2   1999      NYM 97 66
3   2000      NYM 94 68
4   2001      NYM 82 80
5   2002      NYM 75 86
6   2003      NYM 66 95
```