Chick weights

The `ChickWeight` data set is part of the base package `datasets`. See `?ChickWeight` for details on the data. For all of the questions use `dplyr` functions with the pipe `%>%` whenever possible.

```
  library(datasets)
  library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
  library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v stringr 1.4.0
## v tidyr   1.0.0     v forcats 0.4.0
## v readr   1.3.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Submission instructions: Create a folder named `ds202_hw4`, and name the RMarkDown file including your solutions `hw4.Rmd` under this folder. For the submission, create a GitHub repository named `ds202_hw4` under your GitHub account, and push both `hw4.Rmd` and the knitted `hw4.html` before the deadline. I will assume you use the same GitHub username as for your HW3 submission. The latest version of your homework appearing on GitHub before the deadline is graded. *It is crucial to follow the exact folder structure and file names*. This homework is **due on March 4** before class.

1. Get a frequency breakdown of the number of chicks, their average weight and the standard deviation of the weights in each of the diets at the start of the study. Store the resulting data frame in a variable named `weight0` and print.

```
  weight0 <- ChickWeight %>%
    summarize(n = n(), avg_weight = mean(weight), sd_weight = sd(weight))

  print(weight0)
```

```
##     n avg_weight sd_weight
## 1 578   121.8183  71.07196
```

2. Each chick should have twelve weight measurements. Use the `dplyr` package to identify how many measurements are there for each chick. Extract a subset of the data for all chicks with complete information, name the data set `complete`, and print the first 3 rows. (Hint: you might want to use mutate to introduce a helper variable consisting of the number of observations)

3. In the complete data set introduce a new variable that measures the current weight difference compared to day 0. Name this variable `weightgain` and print the first 3 rows of the new data frame.

4. Use `{ggplot2}` to create side-by-side boxplots of `weightgain` by `Diet` for day 21. Change the order of the categories in the Diet variable such that the boxplots are ordered by median `weightgain`, and the

lower limit of the y axis should be 0. Describe the relationship in 2-3 sentences.

5. Answer each of the following questions using numerical and graphical summary. Also explain what you found: (a) Which chick gained the most weight and which gained the least weight? (b) What diets were they on?
6. For the chicks with the least and the most weight at time 0, plot their growth trajectories over time.
7. Add to the previous plot the average growth trajectory for all chicks (calculated as the mean weight for each chick at each time). How do the trajectories compare?
8. What else do you find interesting in the dataset? Show your exploratory analysis.

Note: your submission is supposed to be fully reproducible, i.e. the TA and I will 'knit' your submission in RStudio.