

Homework 2

This homework is due on Jan. 30, 2025 at 11:00pm. Please submit as a pdf file on Canvas.

Problem 1: (6 pts) For this problem you will work with the dataset `txhouse` that has been derived from the `txhousing` dataset provided by `ggplot2`. See here for details of the original dataset: <https://ggplot2.tidyverse.org/reference/txhousing.html>. `txhouse` contains three columns: `city` (listing four Texas cities), `year` (containing four years between 2000 and 2015) and `total_sales` indicating the total number of sales for the specified year and city.

```
txhouse
```

```
# A tibble: 16 × 3
# Groups:   city [4]
  city      year total_sales
  <chr>    <int>     <dbl>
1 Austin   2000     18621
2 Austin   2005     26905
3 Austin   2010     19872
4 Austin   2015     18878
5 Dallas   2000     45446
6 Dallas   2005     59980
7 Dallas   2010     42383
8 Dallas   2015     36735
9 Houston  2000     52459
10 Houston 2005     72800
11 Houston 2010     56807
12 Houston 2015     48109
13 San Antonio 2000     15590
14 San Antonio 2005     24034
15 San Antonio 2010     18449
16 San Antonio 2015     16455
```

Use `ggplot` to make a bar plot of the total housing sales (column `total_sales`) for each year, color the bar borders white, and fill the bars by `city`.

```
# Your code goes here.
```

Problem 2: (8 pts) Modify the plot from Problem 1 by placing bars for different cities side-by-side, rather than stacked. Also, reorder the bars within each year by descending order of total sales. See the slides from the lecture on visualizing amounts for hints on how to do this.

(You do not need to clean up the title of the legend or any of the axis labels. This problem is entirely about bar placement.)

```
# Your code goes here.
```

Problem 3: (6 pts) For this problem you will be working with the `penguins2` dataset, which is a slightly modified version of the `penguins` dataset from the `palmerpenguins` package.

```
penguins2
```

```
# A tibble: 333 × 8
  species island  bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
  <fct>   <fct>          <dbl>         <dbl>          <int>      <int>
1 Adelie  Torgersen         39.1          18.7            181       3750
2 Adelie  Torgersen         39.5          17.4            186       3800
3 Adelie  Torgersen         40.3           18             195       3250
4 Adelie  Torgersen         36.7          19.3            193       3450
5 Adelie  Torgersen         39.3          20.6            190       3650
6 Adelie  Torgersen         38.9          17.8            181       3625
7 Adelie  Torgersen         39.2          19.6            195       4675
8 Adelie  Torgersen         41.1          17.6            182       3200
9 Adelie  Torgersen         38.6          21.2            191       3800
10 Adelie Torgersen         34.6          21.1            198       4400
# i 323 more rows
# i 2 more variables: sex <fct>, year <int>
```

Make a scatter plot of bill length (column `bill_length_mm`) versus body mass (column `body_mass_g`), and color the points by island (column `island`). Use scale functions to adjust the axis title for both axes. For the bill length axis, also adjust the limits from 28 to 72 and place axis ticks at 30, 50, and 70.

```
# Your code goes here.
```