

AE 2: Bike rentals in DC (continued)

Exploring and modeling relationships

Add your name here

Bike rentals in DC

```
library(tidyverse)
library(tidymodels)
library(patchwork)
```

Data

Our dataset contains daily rentals from the Capital Bikeshare in Washington, DC in 2011 and 2012. It was obtained from the `dcbikeshare` data set in the `dsbox` R package.

We will focus on the following variables in the analysis:

- `count`: total bike rentals
- `temp_orig`: Temperature in degrees Celsius
- `season`: 1 - winter, 2 - spring, 3 - summer, 4 - fall

Click [here](#) for the full list of variables and definitions.

```
bikeshare <- read_csv("data/dcbikeshare.csv")
```

See AE 1 for the first part of this analysis.

Daily counts, temperature, and season

Exercise 1

In the raw data, seasons are coded as 1, 2, 3, 4 as numerical values, corresponding to winter, spring, summer, and fall respectively. Recode the `season` variable to make it a categorical variable (a factor) with levels corresponding to season names, making sure that the levels appear in a reasonable order in the variable (i.e., not alphabetical).

```
# add code developed during livecoding here
```

Exercise 2

Next, let's look at how the daily bike rentals differ by season. Let's visualize the distribution of bike rentals by season using density plots. You can think of a density plot as a "smoothed out histogram". Compare and contrast the distributions. Is this what you expected? Why or why not?

```
# add code developed during livecoding here
```

[Add your answer here]

Exercise 3

We want to evaluate whether the relationship between temperature and daily bike rentals is the same for each season. To answer this question, first create a scatter plot of daily bike rentals vs. temperature faceted by season.

```
# add code developed during livecoding here
```

Exercise 4

- Which season appears to have the **strongest** relationship between temperature and daily bike rentals? Why do you think the relationship is strongest in this season?
- Which season appears to have the **weakest** relationship between temperature and daily bike rentals? Why do you think the relationship is weakest in this season?

[Add your answer here]

Modeling

Exercise 5

Filter your data for the season with the strongest apparent relationship between temperature and daily bike rentals.

```
# add code developed during livecoding here
```

Exercise 6

Using the data you filtered in [Exercise 5](#), fit a linear model for predicting daily bike rentals from temperature for this season.

```
# add code developed during livecoding here
```

Synthesis

Exercise 7

Suppose you work for a bike share company in Durham, NC, and they want to predict daily bike rentals in 2022. What is one reason you might recommend they use your analysis for this task? What is one reason you would recommend they not use your analysis for this task?

[Add your answer here]