keeping up with week 5
# Week 4 - Data importing & recoding

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>YouTube</th>
<th>MediaHopper</th>
<th>Slides</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Keeping up with IDS: Week 4</td>
<td><img src="https://i.imgur.com/3.png" alt="YouTube" /></td>
<td><img src="https://i.imgur.com/3.png" alt="MediaHopper" /></td>
<td><img src="https://i.imgur.com/3.png" alt="Slides" /></td>
<td>19:58</td>
</tr>
<tr>
<td>02</td>
<td>Data types</td>
<td><img src="https://i.imgur.com/3.png" alt="YouTube" /></td>
<td><img src="https://i.imgur.com/3.png" alt="MediaHopper" /></td>
<td><img src="https://i.imgur.com/3.png" alt="Slides" /></td>
<td>19:28</td>
</tr>
<tr>
<td>03</td>
<td>Data classes</td>
<td><img src="https://i.imgur.com/3.png" alt="YouTube" /></td>
<td><img src="https://i.imgur.com/3.png" alt="MediaHopper" /></td>
<td><img src="https://i.imgur.com/3.png" alt="Slides" /></td>
<td>30:29</td>
</tr>
<tr>
<td>04</td>
<td>Importing data</td>
<td><img src="https://i.imgur.com/3.png" alt="YouTube" /></td>
<td><img src="https://i.imgur.com/3.png" alt="MediaHopper" /></td>
<td><img src="https://i.imgur.com/3.png" alt="Slides" /></td>
<td>20:22</td>
</tr>
<tr>
<td>05</td>
<td>Recoding data</td>
<td><img src="https://i.imgur.com/3.png" alt="YouTube" /></td>
<td><img src="https://i.imgur.com/3.png" alt="MediaHopper" /></td>
<td><img src="https://i.imgur.com/3.png" alt="Slides" /></td>
<td>17:00</td>
</tr>
<tr>
<td>06</td>
<td>AE: Hotels + Data types</td>
<td><img src="https://i.imgur.com/3.png" alt="YouTube" /></td>
<td><img src="https://i.imgur.com/3.png" alt="MediaHopper" /></td>
<td><img src="https://i.imgur.com/3.png" alt="Slides" /></td>
<td>19:28</td>
</tr>
<tr>
<td>07</td>
<td>AE: Nobels + Sales + Data import</td>
<td><img src="https://i.imgur.com/3.png" alt="YouTube" /></td>
<td><img src="https://i.imgur.com/3.png" alt="MediaHopper" /></td>
<td><img src="https://i.imgur.com/3.png" alt="Slides" /></td>
<td>20:22</td>
</tr>
</tbody>
</table>
Which of the following would **necessarily** improve this visualization? Check all that apply.

- [ ] Replace x-axis labels with month names.
- [ ] Flip the coordinates so months are on the y-axis and mean departure delays are on the x-axis.
- [x] Change the colors of the lines.
- [ ] Indicate units of mean departure time.
- [ ] Change the background color of the plot.
- [ ] Place airline names at the end of lines on the plot instead of in the legend.
- [ ] Add a title.
- [ ] Fix the axis labels to remove underscores and use proper capitalization.
- [ ] Use bars instead of lines to represent the data.
Which of the following would necessarily improve this visualization? Check all that apply.

- Use airline names instead of carrier codes. ✓
- Change the background color of the plot. ✗
- Use a different color for each bar. ✗
- Add a title. ✓
- Place carrier names at an angle so they don’t overlap. ✗
- Use a histogram to represent the data. ✗
- Fix the scales of the axes so that the carriers are in the same order for each and the counts are on the same scale. ✓
- Flip the coordinates so carriers are on the y-axis and counts are on the x-axis. ✓
Two-table verbs

It's rare that a data analyst sees only a single table of data. In practice, you'll normally have many tables that contribute to an analysis, and you need tools to handle them. In `dplyr`, there are three families of verbs that work with two tables at a time:

- **Joining** verbs, which add new variables to one table from another or vice versa.
- **Filtering** verbs, which filter observations from one table based on whether or not they match an observation in the other table.
- **Set operations**, which combine the observations in the two tables as if they were set elements.

The `left_join()` function joins two tables under the assumption that you have `helpdata`, where the rows are observations and the columns are variables. If you're not familiar with that function, it renamed `full_join()` in `dplyr`.

All these verbs work similarly. The first two arguments are `x` and `y`, and provide the tables to combine. The output is always a named data frame with the same types as `x`.
week 5 - Communicating data science results effectively

- Peer evaluations
- Project proposals