## **Course content:**

**Goals for data transformation might be:**

* Data **organization**: better organized data is easier to use
* Data **compatibility**: different applications or systems can then use the same data
* Data **migration**: data with matching formats can be moved from one system to another
* Data **merging**: data with the same organization can be merged together
* Data **enhancement**: data can be displayed with more detailed fields
* Data **comparison**: apples-to-apples comparisons of the data can then be made

| **Wide data is preferred when** | **Long data is preferred when** |
| --- | --- |
| Creating tables and charts with a few variables about each subject | Storing a lot of variables about each subject. For example, 60 years worth of interest rates for each bank |
| Comparing straightforward line graphs | Performing advanced statistical analysis or graphing |

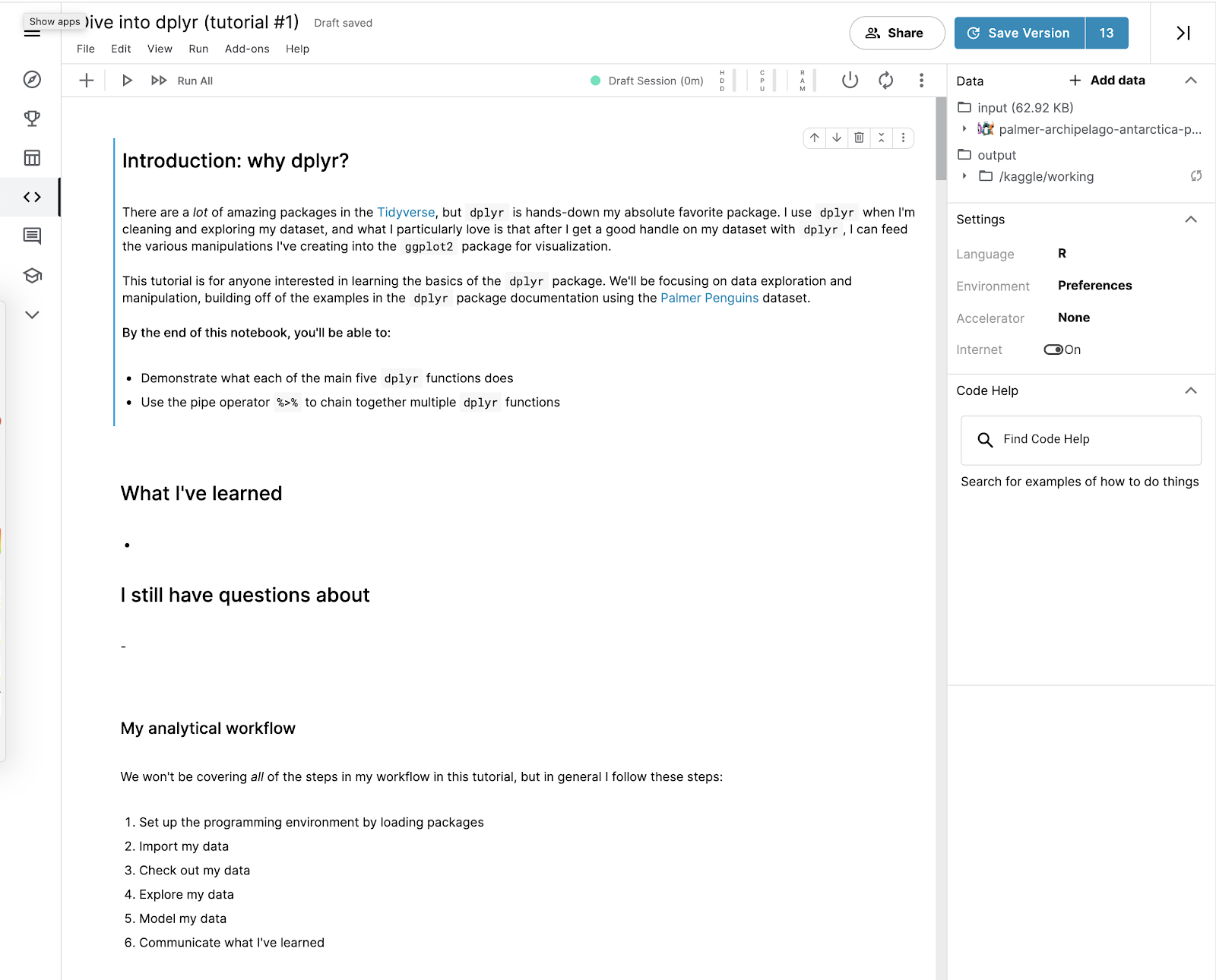
## **Edit a notebook**

****

Now, take a look at a specific notebook: [Dive into dplyr (tutorial #1)](https://www.kaggle.com/jessemostipak/dive-into-dplyr-tutorial-1) by Jesse Mostipak. Follow these steps to learn how to edit notebooks:

1. Click on the link to open up the notebook. It contains the dataset you’ll work with later on.

2. Click on the **Copy and Edit** button at the top-right to make a copy of the notebook in your account. Now, the notebook appears in **Edit** mode. Edit mode lets you make changes to the notebook if you want.



This notebook is private. If you want to share your work, you can choose to make it public. When you copy and edit another Kaggler’s work, always make meaningful changes to the notebook before publishing it. That way, you’re not misrepresenting someone else’s work as your own.

3. Take a moment to explore the Edit mode of the notebook.

Some of this may seem unfamiliar—and that’s just fine. By the end of this course, you’ll know how to create a notebook like this from scratch!

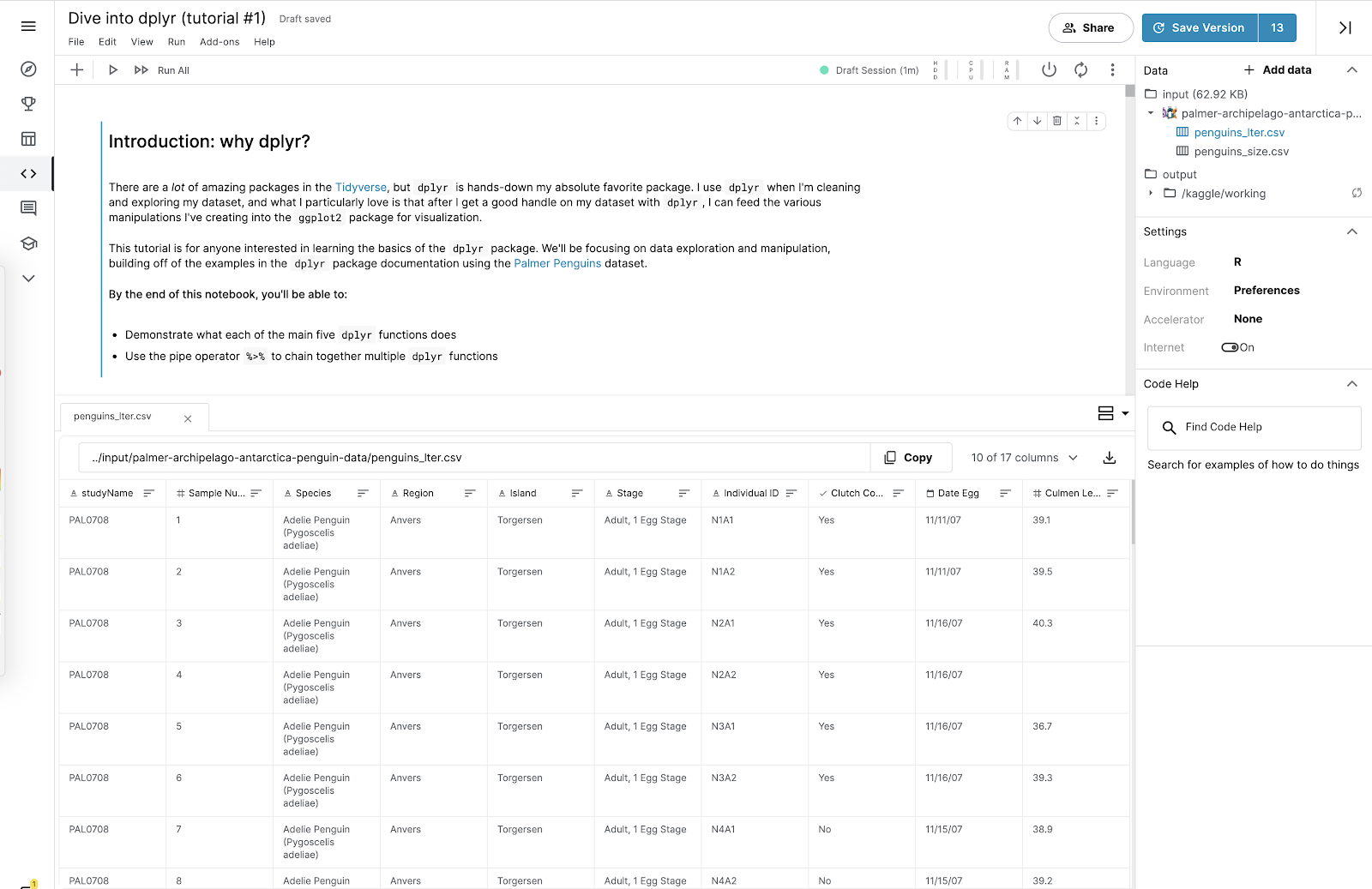
## **Working with datasets in notebooks**

****

Now, you can check out the data!

In this notebook, you’ll find the data in a box labeled **Data** at the top-right of your screen. In the box, there’s an input folder with the title: **palmer-archipelago-antarctica-penguin-data**. Follow these instructions to explore the datasets and learn more about the data within them:

1. Click on this title. Two .csv files appear: **penguins\_lter.csv** and **penguins\_size.csv**. Click on one of them. At the bottom of the notebook, you’ll now find an interactive data table with all the information from the dataset.



2. Click on the other .csv file. This opens a second tab with the second dataset.

3. Take a moment to check out each dataset.

4. Sort the data in each column by clicking on the **horizontal bars** to the right of each column name.

5. Click on the button that says **10 of 17 columns** to change the columns that are visible in the table.

In the dropdown menu, there’s a checkmark next to the name of each column that appears in the table. Checking or unchecking one of these boxes will change what data is presented.

Congratulations! You’ve explored several ways to interact with the dataset. This will help you get familiar with the Kaggle interface. You can save the notebook you worked in for future reference. Coming up, you’ll learn more about other ways you can use Kaggle.