Data-focused Web Applications using Streamlit

An Introduction to Streamlit

Notes Prepared By: Dr. Puneet Agarwal
Email: pagarw05@calpoly.edu
Industrial and Manufacturing Engineering Department
California Polytechnic State University, San Luis Obispo, CA

What is Streamlit?

- ⋆ It is a *web application framework* that helps to build and deploy *Python-based web applications*.
- ⋆ These applications can be used to share *analytics results*, build complex *interactive experiences*, and illustrate new *machine learning models*.
- ⋆ Developing and deploying Streamlit apps is incredibly *fast* and *flexible*.

1 Installing Streamlit

To install streamlit, run the following code in a terminal:

```
pip install streamlit
```

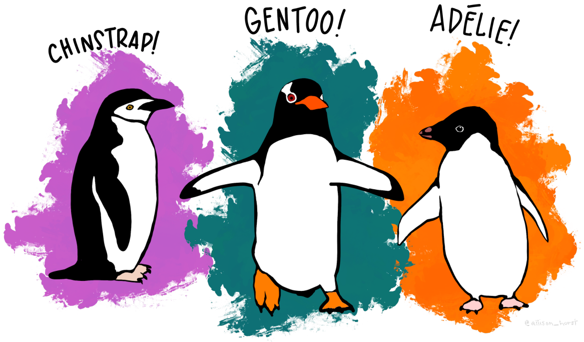
2 Uploading, Downloading, and Manipulating Data

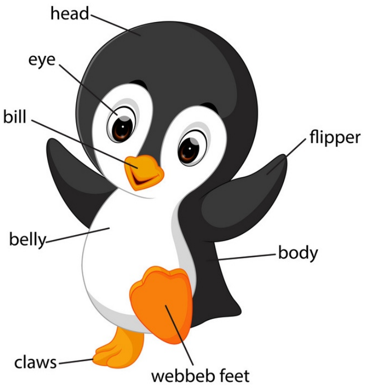
2.1 The Setup - Palmer's Penguins

Dataset: `penguins.csv`

Data is collected and made available by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER, a member of the Long Term Ecological Research Network.

About the dataset: It includes data on 344 individual penguins with 3 species represented: **Adelie**, **Chinstrap**, **Gentoo**. It displays the measurements of various body features of a penguin.





First, let's create a folder for our new app in our existing `Streamlit Applications` folder using the following code block:

```
mkdir penguin_app    # Create a folder in the current directory
cd penguin_app       # Change your directory to this new folder
touch penguins.py    # Create your app (Python) file
```

After this, download the `penguins.csv` file and put it in the `penguin_app` folder. Now, our folder should have the `penguins.py` file and `penguins.csv` file.

2.2 Displaying First 5 Rows of Dataset

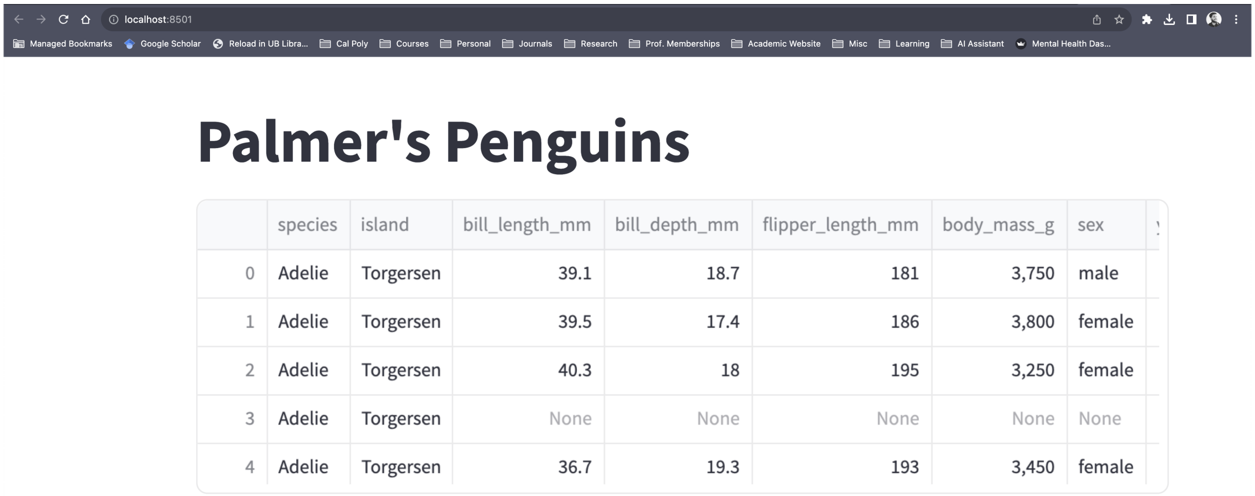
Use the following code:

```
# Import packages
import streamlit as st
import pandas as pd

# Create a title for our app
st.title("Palmer's Penguins")

# Import data
penguins_df = pd.read_csv("penguins.csv")

# Display rows using st.write() function
st.write(penguins_df.head())
```

Now we can test the above code by running the following in the terminal:

```
streamlit run penguins.py
```

The following figure is a snapshot of our app in a browser:

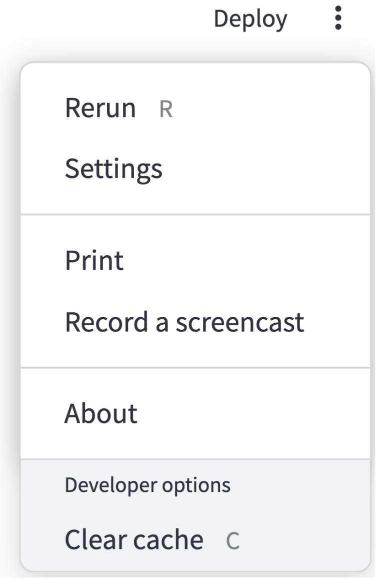


Palmer's Penguins

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	y
0	Adelie	Torgersen	39.1	18.7	181	3,750	male	
1	Adelie	Torgersen	39.5	17.4	186	3,800	female	
2	Adelie	Torgersen	40.3	18	195	3,250	female	
3	Adelie	Torgersen	None	None	None	None	None	
4	Adelie	Torgersen	36.7	19.3	193	3,450	female	

There are 3 items to note here:

- ⋆ First, we are able to display first five rows of our dataset, along with a title.
- ⋆ Next, we see that the URL points to **localhost:8501**. This tells us that our app is hosted locally (it is not on the internet anywhere) through port 8501.
- ⋆ Third item is the hamburger icon (⋮) at the top right. The following screenshot shows us what happens when we click the icon:



This is the default options panel for Streamlit apps. In this course, we will discuss each of these options in depth. All we have to know for now is that if we want to rerun the app or find settings or the documentation, we can use this icon to find almost whatever we need.

Exploring Palmer’s Penguins

App 1: Scatterplots using User Inputs

The goal of this app is to ask the user to specify one of the species of penguins and to choose two variables to use for a scatterplot.

We will use `st.selectbox()` function to ask the user to select one option from multiple options. We will use this to get the three inputs for our scatterplot:

```python
# Import packages
import streamlit as st
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Main title of the app
st.title("Palmer’s Penguins")

# Our subtitle
# NOTE: st.markdown() enables the use of Markdown, a markup language
# especially useful for writing math equations
st.markdown("Use this Streamlit app to make your own scatterplot about penguins!")

# Select box for species
selected_species = st.selectbox(
    "What species would you like to visualize?",  # First comes the message/question
    ["Adelie", "Gentoo", "Chinstrap"],             # Then the options within a list
)

# Select box for x_var
selected_x_var = st.selectbox(
    "What do you want the x variable to be?",
    ["bill_length_mm", "bill_depth_mm", "flipper_length_mm", "body_mass_g"],
)

# Select box for y_var
selected_y_var = st.selectbox(
    "What about the y?",
    ["bill_depth_mm", "bill_length_mm", "flipper_length_mm", "body_mass_g"],
)
```

Now that we have the `selected_species` variable, we can filter our DataFrame and make a quick scatterplot using the selected `x` and `y` variables:

```
# Import data
penguins_df = pd.read_csv("penguins.csv")

# Filter by species
penguins_df = penguins_df[penguins_df["species"] == selected_species]

# Create scatterplot
fig, ax = plt.subplots()
ax = sns.scatterplot(x = penguins_df[selected_x_var], y = penguins_df[selected_y_var])
plt.xlabel(selected_x_var)
plt.ylabel(selected_y_var)
plt.title("Scatterplot of " + selected_species + " Penguins")
st.pyplot(fig)
```

NOTE: `st.pyplot()` allows us to use `matplotlib` library and pushes our `matplotlib` graph to Streamlit.

Now we can test the above code by running the following in the terminal:

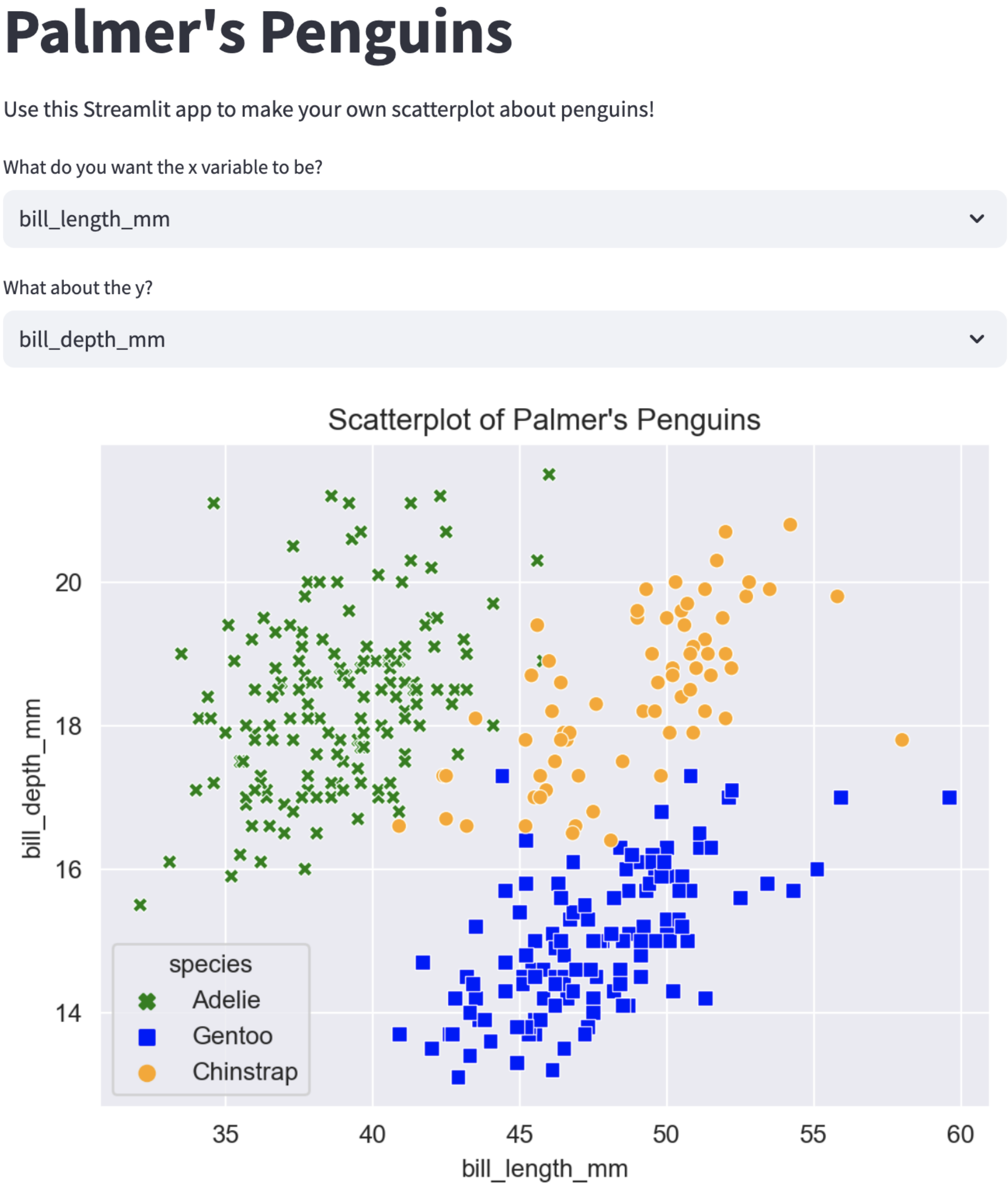
```
streamlit run app.py
```

Replace `app.py` with the name of your Streamlit script.

In-class Practice

Create a scatterplot showing each species using distinct marker shapes and sizes. For instructions, refer to the `seaborn.scatterplot` documentation available at this link.

Your output should resemble something like this:



Palmer's Penguins

Use this Streamlit app to make your own scatterplot about penguins!

What do you want the x variable to be?

bill_length_mm

What about the y?

bill_depth_mm