**Proposal Team 9**

**Objective:**

The purpose of this project is to determine the appropriate programming languages and tools for developers based off certain demographics such as experience and areas of expertise.

**Problem Statement:**

As new programmers or experienced veterans of older languages, developers may find the prospect of choosing a new language or tool complicated or confusing. The data we use should hopefully give them some indication of where various aspects of the industry are focused or at least point them towards the technologies which are more likely to receive aid when problems occur. We are getting the data from the Stack Overflow Annual Developer Survey at : <https://insights.stackoverflow.com/survey/>

There are a variety of reasons why people should care about this topic. Students want to discover popular skills to find jobs, and the career satisfaction of developers in certain fields will help direct this search. Developers may want to understand which tools are most used by others in their industry if they are discontent with their own, and even companies could take language popularity into consideration when starting projects if they care more about ease of hiring over the use of specific tools.

**Approach to Applying Data:**

This survey data consists of many details like employment status, major, or years of programming experience that can be used to group developers into various demographics. The data comes from a major programming community with 8 million users, so it should be relevant to most other programmers. In order to use this data, we will extract the data from the survey information we deem relevant either as useful identifiers for a group of developers or as interesting and useful traits that we can compare among groups. We will utilize Python to extract and transform the data to get these variables which can be used for analysis and visualization of prominent developer traits in various developer groupings. We plan to use the Numpy, Pandas, and Matplotlib modules in Python. In particular, we will likely be using Pandas to extract and format the data relevant to our objective. We chose them initially over other modules because they are covered in detail in class, so we will have some baseline of knowledge. In addition, they are also very popular, proven and well-supported which means references are available from various resources. In terms of functionality, both Numpy and Pandas provides various means of mathematical computation through specialized data structures and Matplotlib is able to provide data visualization capabilities.

**Milestones** Time to reach Members:

Downloading data Nov 14 Chen

Making tables Nov 18 Liu

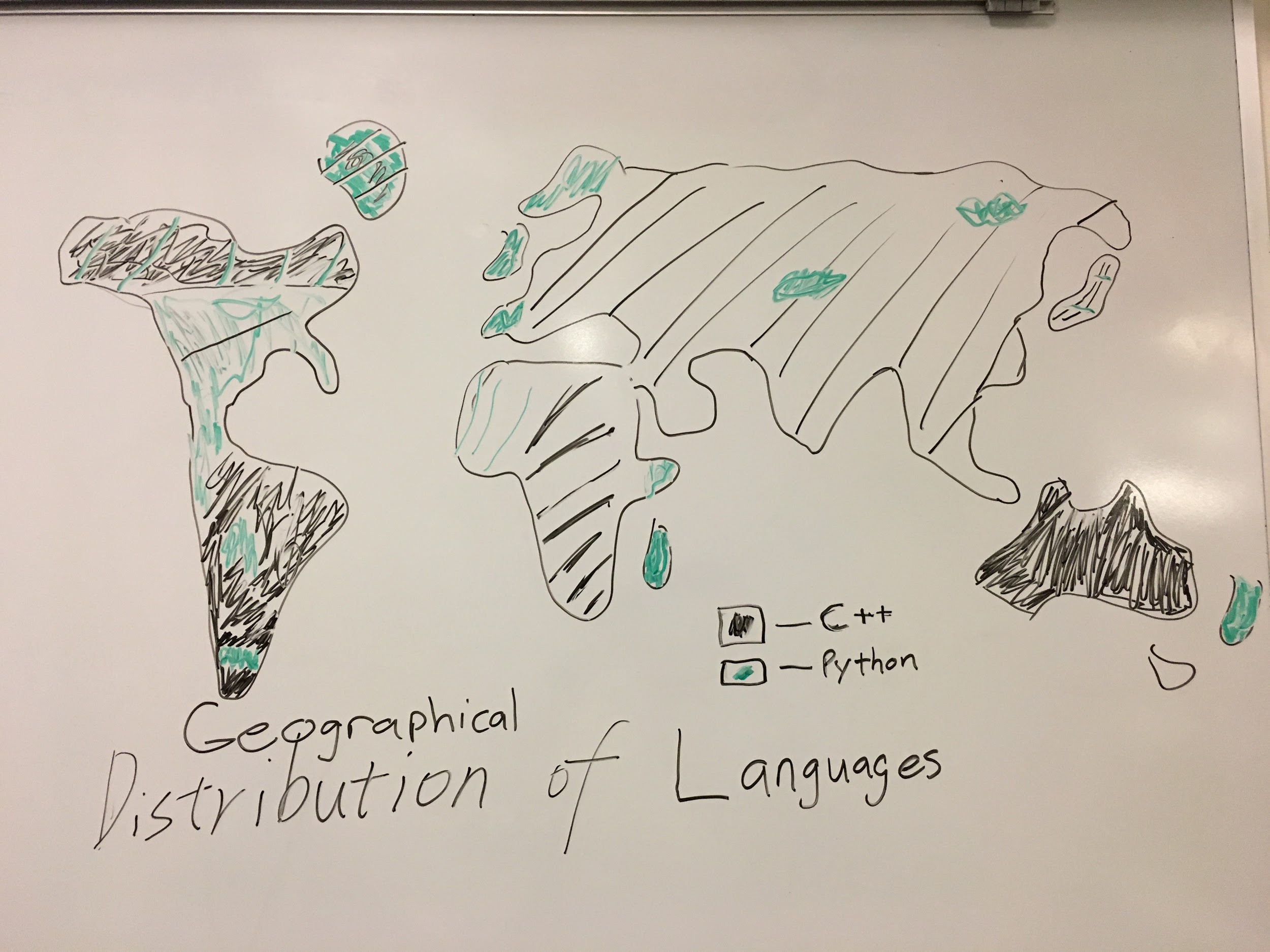
Formatting data Nov 20 Li

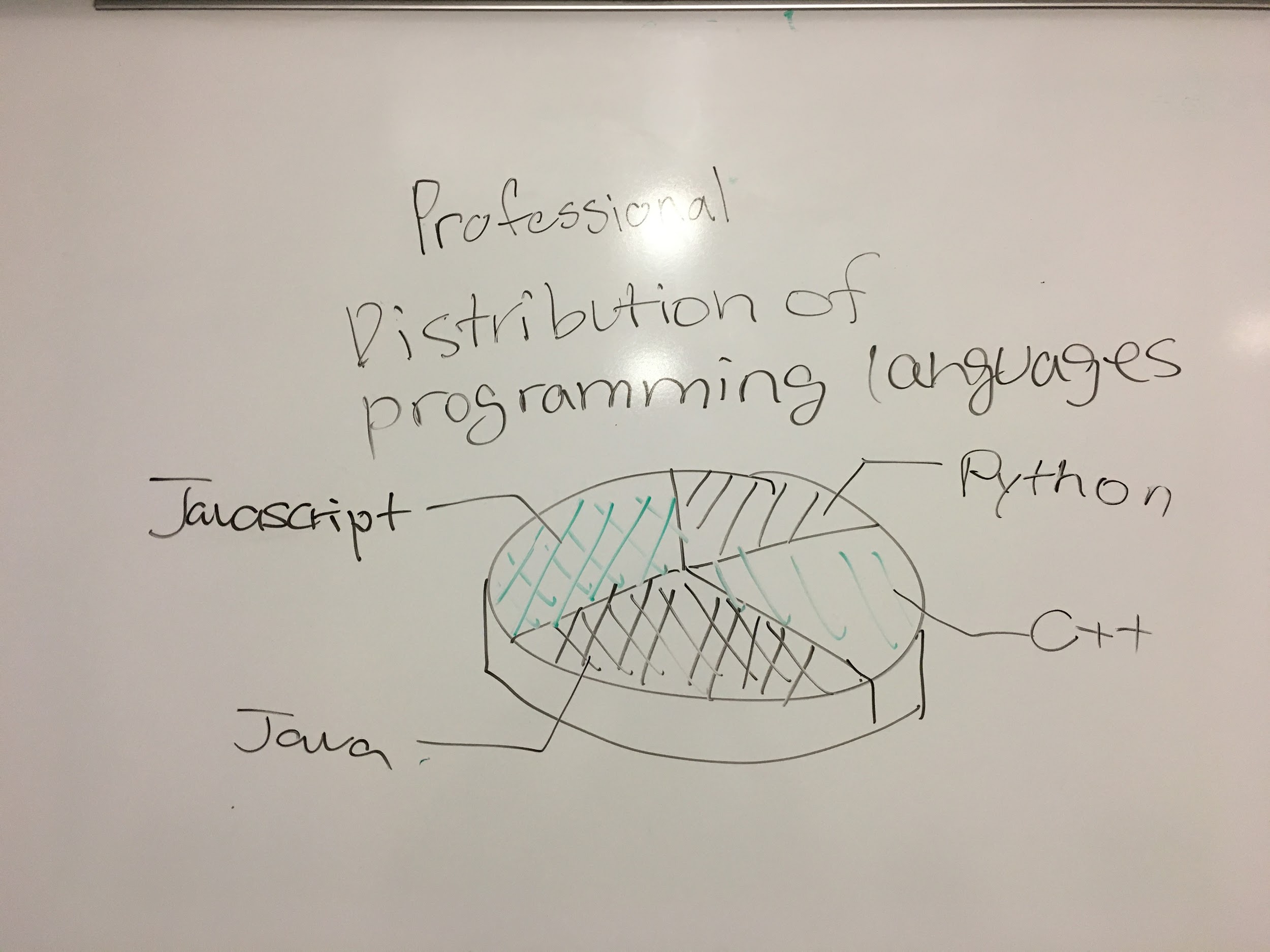
Data Computation and Analysis Nov 27 Pham, Chen, Lin

Visualization of the results Dec 1 Liu, Li

Removing errors Dec 4 Lin, Pham

**Draft Graphics:** There were only two colors in the library so it is much more limited than the real thing.

We could have geographical distributions of various data like programming languages, with gradients or colors impacted by the concentration of different languages in a country. A similar graphic could be utilized for tools or IDEs used around the world, types of developers, or perhaps salaries if there is enough data.



A graphic we could have is one showing the most used languages for a certain type of user, in this case professionals. This could be contrasted with the language distribution among students or be more specific by delving into specific roles like “Web Developer”.