Privacy Terms Final Project Proposal

Group Members:
Ishika Keswani
Group Role: Task Manager. Ishika is the brains of this proposal, as the proposed project is a spinoff from her 10-week SURP project which she presented a poster on. Therefore, she will be responsible for delegating tasks based on both our's relative expertise.
Project Role: Ishika conducted many user studies, and is familiar with common pain points and main misunderstood terms in both IOS and Android Privacy Labels. Ishika will be the mainly responsible person for creating a simplified version of the super complex Privacy Terms. Ishika's insight and intuition will come in super handy for this task. Ishika will also be responsible for leading the web-scraping/online data processing process.

Ulas Ayyilmaz
Group Role: Project Manager: Ulas will be on top of each meeting scheduling, as well as checking if goals are met, upholding a clear structure to achieve the end product in the desired time.
Project Role: Ulas will be responsible for building the backend as well as the front end of an extension that will allow the feature of hovering your mouse on a term and seeing it's simplified version. Ulas will also contribute to webscraping.

Ulas and Ishika will code both "What percent of similar apps found on google play store" feature, as well as the role of demographic background of an individual on their understanding of the terms (data collected by Ishika's research group.

TITLE: Understanding User Privacy in Android Devices

Broad Topic: Exploring terms found in Android Privacy 'Nutrition Labels' and creating an interface/website/app wherein users could input an android app store link (no privacy issues with android which makes web-scraping a much easier process J) and have as an output a simplified version of the terms and whether they are collected or shared, and what percentage of 'similar apps' found on the google play store collect/share each particular variable type. Additionally, if time permits, we could include a feature wherein hovering the cursor over a term would lead to a small projection on the screen containing a simplified explanation/definition of the term to increase accessibility of this information and negate the need to go to a whole different page or look up the definition.

Data & Variables & End Product: Taking inspiration from the New York Times tree in the Decision Trees topic notes on the website, I thought it would be interesting to create a scrollytelling graphic/image of a similar tree which could explore:
1. How **demographics** impact which terms from the android list of **terms** are more or

less widely understood (information available via the full study we conducted as surveyed individuals were required to report their demographics) or how demographics determine whether an individual would get the definition of a term correct, partially correct, wrong, restate the term, or say they don't know the answer.

2. We divided the terms into categories based on the results of a follow up study we conducted wherein users were given the definition/meaning of a term in colloquial language and were asked to suggest phrases/term names for each of them such that it would be easiest to understand. The categories were 1) **Terms whose diction needed to be clarified**, 2) **Terms that were too broad or needed to be split up** into multiple smaller terms, 3) **Terms including the word 'other'**, 4) **Terms with no good alternatives**, 5) **Terms lacking nuance**. We could also see how **correctness** of the term differed across these categories and whether we could use correctness and user demographics to predict which category a term would fall under.

Another use of data in this project could link back to the segment about the 'similar apps' on the app store. We could also use **graphics** and other available data (such as **number of downloads** of an app, **rating of an app**, number of apps with the same function, etc.) to determine whether we can create a model - possibly integrating methods that our out of the scope of Compstats such as deep NNs - to **predict if an app collects/shares more or less data** than other similar apps which perform the same function.