

Final Project Proposal

- Basic Information:

Project Title: Introduction to Whisky for New Whisky Enthusiastic

Name: Maggie Lu, Jacques Sham

Email Address: mlu13@dons.usfca.edu, jjsham@dons.usfca.edu

UIDs: jjsham, 20445531

Mlu13, 20333276

Repository: <https://github.com/jacquessham/USF-CS360-FinalProject>

- Background and Motivation: Have you ever ran into BevMo or Total Wine whisky shelf and confused which bottle of whisky to buy? We love whisky and would like to promote whisky to everyone. However, there are a lot of whisky in the world that new whisky drinkers are confused what to choose. Therefore, we would like to gather the data on whisky and visualize it for the convenience of readers to choose which bottle of whisky is for them.
- Project Objective: New whisky drinkers are always confused with the shelves of whisky and ask: Which bottle is best for me? In this project, we will introduce basic knowledges of whisky to new drinkers, such as the differences among different origins. Therefore, new drinkers will learn what to be noticed when picking up a bottle of whisky. We would like to help them to analyze the data without having they do too much research. So new drinkers have a easier time to understand and choose whisky from the shelf based on their preference. Also, it would help the industry to promote their products.
- Data: The data, whisky.csv, is from SelfBuilt's whiskyanalysis.com: <https://whiskyanalysis.com/> The data consists majority of the famous whisky with their reviewerr's score, profile including favor, type and origin country. The reviews are mostly non-professional whisky tasters.
- Data Processing: whisky.csv consists the following columns:

whisky	Meta Critic	STDEV	#	costs	class	super-cluster	cluster	country	type
--------	-------------	-------	---	-------	-------	---------------	---------	---------	------

Whisky: The name of the whisky and vintage. Some of the whisky does not have vintage due to some distilleries blend different vintage whisky, so we are going to keep the vintage along the name and not going to separate into two columns.

Meta Critic: The mean of normalized score from reviews who have review this whisky.

STDEV: The standard deviation of the Meta Critic.

#: Number of reviewers

Cost: The average retail price range, the following is the definition

\$ <\$30 CAD
\$\$ \$30-50 CAD
\$\$\$ \$50-70 CAD
\$\$\$\$ \$70-125 CAD
\$\$\$\$\$ \$125-300 CAD
\$\$\$\$\$+ >\$300 CAD

Class: Identify the major characteristic of the whisky, includes:

- 1) Single-Malt-Like, Similar to classic Scotch single malt whisky
- 2) Scotch-Like, Similar to classic Scotch blended whisky
- 3) Rye-Like, Similar to classic Canadian or American rye flvors
- 4) Bourbon-Like, Similar to classic American corn-based whisky style

Supercluster: Grouping of Clusters where the characteristics are similar enough to overlap considerably on the principal component analysis. That means the major character of that whisky. Example: Ledaig 42yo Dugadh's Super Cluster is "ABC", while the Cluster is "C" that means the character of Ledaig consits of Identifier "A", "B", and "C", but the outstanding character is "C". We will define the identifier after Cluster.

Cluster: The revised cluster analysis based on Wishart Analysis. The following is the definition of the identifier:

A – Full-bodied, sweet, pronounced sherry – with fruity, honey and spicy notes

B – Full-bodied, sweet, pronounced sherry – with fruity, floral and malty notes, some honey and spicy notes may be evident

C – Full-bodied, sweet, pronounced sherry – with fruity, floral, nutty, and spicy notes, some smoky notes may be evident

E – Medium-bodied, medium-sweet – with fruity, honey, malty and winey notes, some smoky and spicy notes may be evident

F – Full-bodied, sweet and malty – with fruity, spicy, and smoky notes
 G – Light-bodied, sweet, apéritif-style – with honey, floral, fruity and spicy notes, but rarely any smoky notes
 H – Very light-bodied, sweet, apéritif-style – with malty, fruity and floral notes
 I – Medium-bodied, medium-sweet, quite smoky – with some medicinal notes and spicy, fruity and nutty notes
 J – Full-bodied, dry, very smoky, pungent – with medicinal notes and some spicy, malty and fruity notes
 R0 – ($r=0\%$) – “No Rye” whisky with 0 rye grain
 R1 – ($r\leq 10\%$) – “Low Rye” whisky of 10% or less rye grain
 R2 – ($10\% < r \leq 15\%$) – “Standard Rye” whisky of 10-15% rye grain
 R3 – ($15\% < r < 51\%$) – “High Rye” whisky of more 15% rye
 R4 – ($r \geq 51\%$) – “Rye” whisky of more than 51% rye

Country: Country of Origin of the whisky.

- Cleanup: We would like to modified some contents within the dataset. Meta Critic will be replaced by “score”.
- # sign will be substituted by “num_viewers” which stands for Number of Viewer. We both agreed that STDEV, which stands for standard deviation, is not necessary needed for this project, so we decide to remove column STDEV.
- We decide to remove Supercluster and keep Cluster since Cluster listed the most outstanding character of certain whisky. We believe the most outstanding character could be one of the most important considerations for new whisky drinkers.
- For column Cost, we will convert \$ into number scale, which the scale will be from 1 to 6. For example. \$ will be converted to 1 and \$\$\$\$\$+ will be converted to 6. The reason of converting \$ is to avoid any problem when reading in data. We will use R language for all modifications. R code will be provided.
- Visualization Design
 - General Ideas: At the beginning, we would like to introduce background and history of whisky for different region. In the following part, present our analyzation to new drinkers. The article ends with recommendation by user’s preference.
 - Display data through:
 - Bar chart
 - Pie chart
 - Geomap
 - Packed Bubble

■ Table (For Recommendation to user)

○ Five Design Sheet

<p>1. Ideas</p> <ul style="list-style-type: none">• Geomap to display origin \Rightarrow background• Bar chart - compare average between different origins• Pie chart - Explain the database by user's selection of:<ul style="list-style-type: none">- Origin of country- Character- Class- Type• story - history of whisky<ul style="list-style-type: none">- history of whisky in major origins- explain each characteristic• recommendation based on user's preference• Scatterplot based on Origin of country and score• Randomly generated recommendation for user	<p>2. Filter</p> <ul style="list-style-type: none">- Scatterplot is same as bar chart's content, take out- Randomly generated recommendation
<p>4. Combine & Refine</p> <ul style="list-style-type: none">• Bar chart - when user click once, show a baseline to demonstrate country above global average score• Packed Bubble/Pie chart - country with more vintage has larger radius for bubble	<p>3. Categorize</p> <ol style="list-style-type: none">1) Geomap - World map \hookrightarrow show each origin2) Bar chart - Average score3) Pie chart - Data explanation4) Text - Story5) Filter - Filter out what user don't like <p>5. Questions</p> <ul style="list-style-type: none">- recommendation- basic information of whisky/history- analyze for user

Layout

Title

History/Story of Whisky

Top 5 countries of origins \Rightarrow introduction

pie chart

Database explanation

bar chart

comparison \Rightarrow score

Geomap

Title: Draft 1

Author: Maggie, Jacques

Date: 4/5/2018

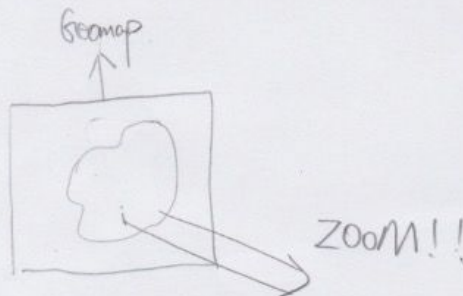
Sheet: 1

Task:

Operations

- Pie chart: gives a overview of our database \Rightarrow explains each category
- Bar chart: show average score for each country.
- Click function: for user to zoom in & receive more details in Geomap.

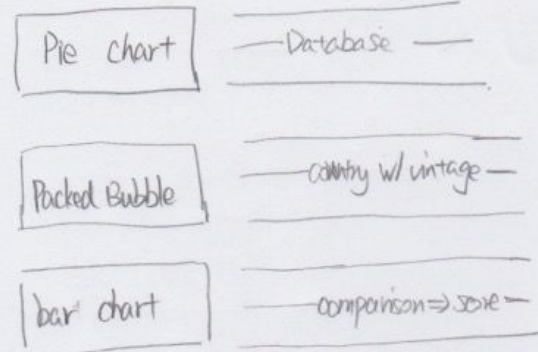
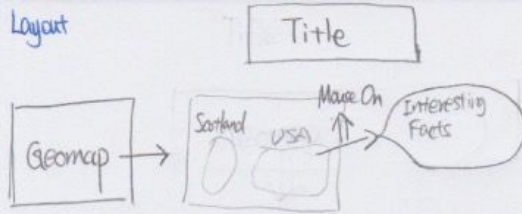
Focus



Discussion

- Chart: clear layout, follow the principle of storytelling. (+)
- Geomap: when user zoom into each country, no useful analyzing/info for new drinkers. (-)
- Too much text in the ~~beginning~~ beginning. (-)

Layout



Title: Draft 2.

Author: Jacques, Maggie.

Date: 4/5/2018

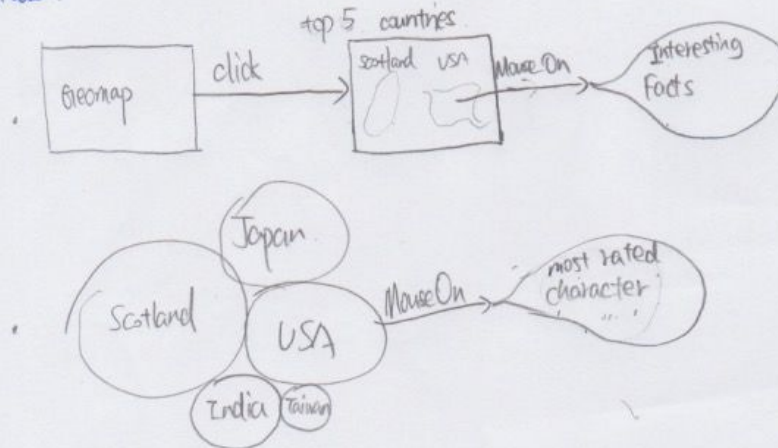
Sheet: 2

Task:

Operations.

- Geomap: show interesting facts for top 5 countries through interaction
- Packed Bubble: show most rate cluster of the country

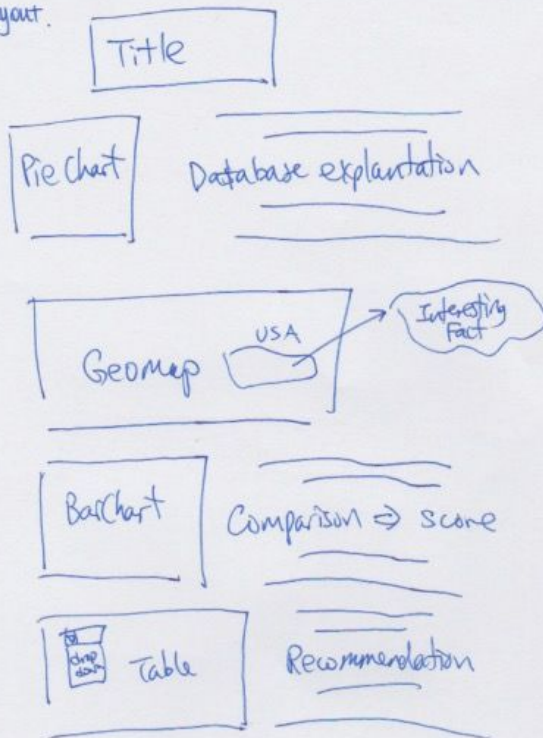
Focus.



Discussion.

- Bar/Pie chart: clear layout, follow, follow principle (+)
- Geomap: show stories through interaction (+)
- Bubble chart: not enough general info for each country (-).

Layout.



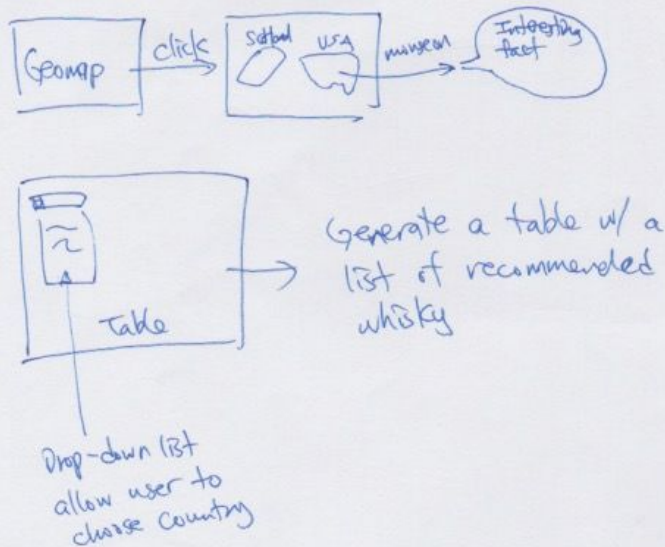
Title: Draft 3
 Author: Jacques, Maggie
 Date: 24/05/2018
 Sheet: 24

Task:

Operations:

- Geomap: Interesting fact for top 5 countries through interaction
- Table: Filter the selection through drop-down list

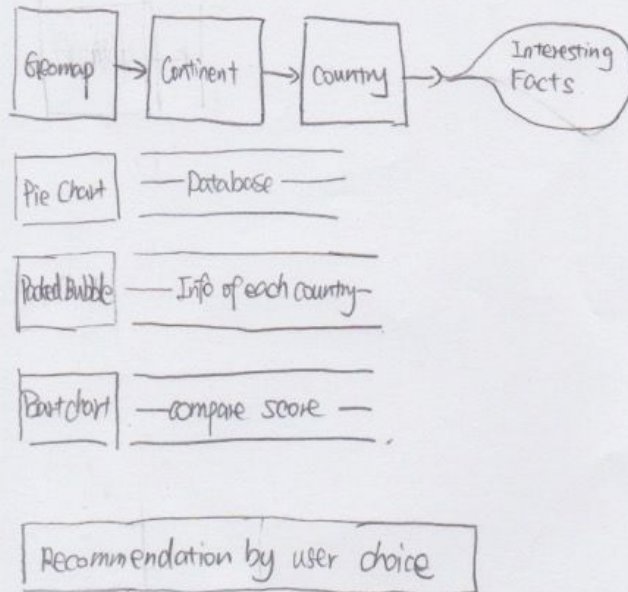
Focus.



Discussion

- Not too much Overview before getting into detail (-)
- User can understand more on our background first (+)
- User can get a list of his/her favorite whisky (+)

Layout



Title: Final draft

Author: Jacques, Maggie.

Date: 4/5/2018.

Sheet: 5.

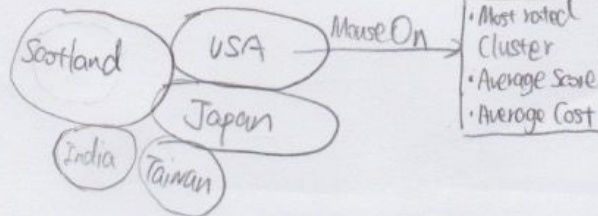
Task:

Operations

- dropdown list
- pie chart - click (character view)
- zoom in
- Mouse On

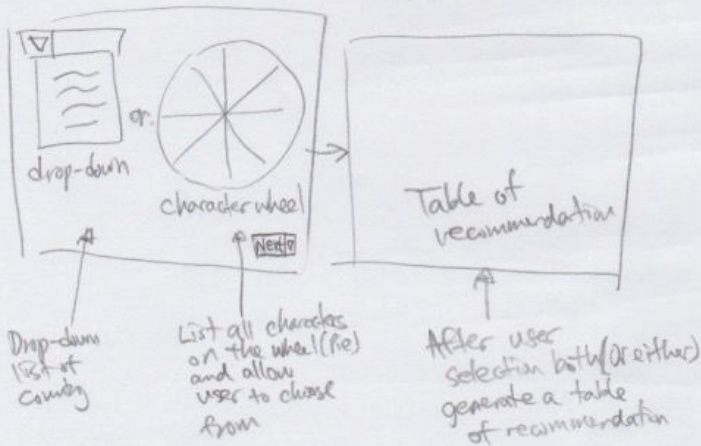
Focus

Packed Bubble:



Detail

- Details on demand (+)
- Visualize the contrast of characters on the "character wheel" (+)
- Too much interaction (-)



In the final decision, we decided to take the final draft to be our prototype.

Must-Have Features

- Geomap to display the whisky region for overview
- Interesting facts on whisky regions
- Details on demand by pie chart/bar chart to explain the database
- Filter function for recommending whisky for readers
- Packed bubble chart
- Special symbols to represent element such as cost.

Optional Features

- Use singled colored ellipse to represent single-malt & mixed colored ellipse to represent blend malt
- Either use dropdown list or pie chart with clicked function for user's selection

- Project Schedule

- Revised Proposal: 04/09/2018
- Create a bar chart: 04/11/2018
- Research interesting facts on whisky by countries : 04/11/2018
- Finish all the remaining graphs: 04/14/2018
- Put all the graphs onto html: 04/15/2018
- Alpha Release: 04/16/2018
- Revise html: 04/23/2018
- Beta Release: 4/25/2018
- Project Presentation: 05/07/2018 or 05/09/2018