Final Term Project Guideline

Session 1: 5:00– 6:30 pm, October 27 (Wed), 2021
Session 2: 5:00– 7:30 pm, November 3 (Wed), 2021
Session 3: 5:00– 6:30 pm, November 10 (Wed), 2021

Presentation plan overview

- 5 min presentation.
- a short Q&A session (2 mins).

It is designed to help you adjust your project plan as you progress further into the semester and better understand the problems and opportunities.

1. Goal(s) and overview of your project – Invent a project name or your dashboard name.
2. List examples of problems and examples of analyses you plan to support through the data visualization and through your dashboard
3. Identify and list data sources.

Name	Presentation
Devin M	10/27/2021
Zhixia D	10/27/2021
Chenyu Z	10/27/2021
Dongping Z	10/27/2021
Honglin	10/27/2021
Peter Z	10/27/2021
Anthony L	11/3/2021
Allison G	11/3/2021
Jackie R	11/3/2021
Bala S	11/3/2021
Jack G	11/3/2021
Lenny A	11/3/2021
Nidhi S	11/3/2021
Peng C	11/3/2021
Shariq K	11/3/2021
Sybille L	11/3/2021
Shubhangi	11/3/2021
William M	11/3/2021
Alexander Z	11/3/2021
Vandana G	11/3/2021
Lynn P	11/3/2021
Vidyalaxmi K	11/10/2021
Ankit B	11/10/2021
Piyush S	11/10/2021
Graham R	11/10/2021
Huiqun H	11/10/2021
Kaustubh P	11/10/2021
Majid F	11/10/2021
Michael M	11/10/2021
Seth F	11/10/2021
Sanjiv D	11/10/2021
Frank Z	11/10/2021

Final Term Project Guideline

Session 1: 5:00– 6:30 pm, October 27 (Wed), 2021
Session 2: 5:00– 7:30 pm, November 3 (Wed), 2021
Session 3: 5:00– 6:30 pm, November 10 (Wed), 2021

Include some look and feel of your Dashboard visualization – seeing is believing!

1. You are welcome to include some **mockup visualizations** in your presentation (e.g., on PPT). You can delineate which ones would be easy to accomplish and which may be more challenging.
2. You conclude your presentation by giving a short live demo of **your prototype Dashboard system** which you will be developing in Part 1 of Homework 6/7 and even an extended version using your project data.

Each of you should prepare a visual aid (e.g., PPT) for your presentation which can include items described in the above.

Peer Evaluation form which should be filled and submitted to HuskyCT after both sessions are over. Participation credits are assigned for the submission.

To minimize the transition time between presenters, speakers should have the sharable materials ready for Collaborator sharing and should practice how to share these from your computing environment. Help materials are given.

* Recording must not be used outside of the UCONN learning environment.

Name	Presentation
Devin M	10/27/2021
Zhixia D	10/27/2021
Chenyu Z	10/27/2021
Dongping Z	10/27/2021
Honglin	10/27/2021
Peter Z	10/27/2021
Anthony L	11/3/2021
Allison G	11/3/2021
Jackie R	11/3/2021
Bala S	11/3/2021
Jack G	11/3/2021
Lenny A	11/3/2021
Nidhi S	11/3/2021
Peng C	11/3/2021
Shariq K	11/3/2021
Sybille L	11/3/2021
Shubhangi	11/3/2021
William M	11/3/2021
Alexander Z	11/3/2021
Vandana G	11/3/2021
Lynn P	11/3/2021
Vidyalaxmi K	11/10/2021
Ankit B	11/10/2021
Piyush S	11/10/2021
Graham R	11/10/2021
Huiqun H	11/10/2021
Kaustubh P	11/10/2021
Majid F	11/10/2021
Michael M	11/10/2021
Seth F	11/10/2021
Sanjiv D	11/10/2021
Frank Z	11/10/2021

Key to Success in Term Project - Modeling

What problems to solve? – You should have one or more themes that string together the visualizations that you offer in your Dashboard. What you are trying to communicate?

Hypothesis testing

null hypothesis
alternative hypothesis

⇨ Reject the null hypothesis in favor of the alternative or not? ⇨

t- Test
p-value
ROC
FDR

Examples

Outfielders are better hitters.
Player position has nothing to do with hit ratio.

COVID-19 is a man's disease.
Gender has nothing to do with COVID-19.

Apple's stock prices do not follow Dow Jones Index (time series data).
There is no difference between Apple's stock trajectory and Dow Jones index trajectory.

Asian countries make more horror movies.
There is no country/region difference as far as movie content is concerned.

Showing plots only for the sake of plot may not be interesting.

Dataset Topic Determined
HIV GSE PrEP
Covid-19/Kaggle
Alzheimer's Disease
Liver Gene Expression
Rain in Australia Kaggle
Water quality
COVID-19 GSE
Twitch.tv
Covid-19/Kaggle
Netflix
U.S. pollution Data
Heart Failure Clinical/Kaggle
DS Job Market/Kaggle
fMRI datasets
Apple Business
Musicnet
Cancer Data
Tweets
Hardware simulator logs
DC Bike Sharing
MAHNOB
MLB Hitting/Pitching
Health statistics

Key to Success in Term Project - Modeling

What problems to solve? – Would the following analysis meaningful for your dataset?

Correlation Analysis

Linear regression
Pearson correlation

Example

Contamination of NO2 is highly correlated with contamination of SO2 in US cities.

Pay scale and hit ratio are highly correlated among MLB players.

What aspects of DC life may influence DC bike sharing?

Dataset Topic Determined
HIV GSE PrEP
Covid-19/Kaggle
Alzheimer's Disease
Liver Gene Expression
Rain in Australia Kaggle
Water quality
COVID-19 GSE
Twitch.tv
Covid-19/Kaggle
Netflix
U.S. pollution Data
Heart Failure Clinical/Kaggle
DS Job Market/Kaggle
fMRI datasets
Apple Business
Musicnet
Cancer Data
Tweets
Hardware simulator logs
DC Bike Sharing
MAHNOB
MLB Hitting/Pitching
Health statistics

Key to Success in Term Project - Modeling

What problems to solve? – Would the following analysis meaningful for your dataset?

Network Analysis

Directional
Non-directional

Example

Do tweeters form a network of subnetwork (i.e., birds of a feather)?

Can the tweeters be clustered?

Can gene expression relationships be constructed into a directional network?

Dataset Topic Determined
HIV GSE PrEP
Covid-19/Kaggle
Alzheimer's Disease
Liver Gene Expression
Rain in Australia Kaggle
Water quality
COVID-19 GSE
Twitch.tv
Covid-19/Kaggle
Netflix
U.S. pollution Data
Heart Failure Clinical/Kaggle
DS Job Market/Kaggle
fMRI datasets
Apple Business
Musicnet
Cancer Data
Tweets
Hardware simulator logs
DC Bike Sharing
MAHNOB
MLB Hitting/Pitching
Health statistics

Key to Success in Term Project - Modeling

What problems to solve? – Would the following analysis meaningful for your dataset?

Clustering/Classification

K-means
Gaussian mixed model

Example

Can heart failure cases be grouped?.

How can cancer patients be grouped?

Can tweets be grouped by various features? Even by content?

How can US Data Scientist Jobs be grouped and/or classified?

Dataset Topic Determined
HIV GSE PrEP
Covid-19/Kaggle
Alzheimer's Disease
Liver Gene Expression
Rain in Australia Kaggle
Water quality
COVID-19 GSE
Twitch.tv
Covid-19/Kaggle
Netflix
U.S. pollution Data
Heart Failure Clinical/Kaggle
DS Job Market/Kaggle
fMRI datasets
Apple Business
Musicnet
Cancer Data
Tweets
Hardware simulator logs
DC Bike Sharing
MAHNOB
MLB Hitting/Pitching
Health statistics

Key to Success in Term Project - Modeling

What problems to solve? – Would the following analysis meaningful for your dataset?

Geospatial / 2D Map data visualization

Geocoded
Hotspot

Example

Show heatmap of NO2 (and others) contamination over US cities/regions.

Show a baseball positional map mapped to aggregated hit ratios (outfielders hit better).

Show distribution of Apple sales over the world (countries loving Apple products).

Contrast biological functional map vs. differential gene expression (does it show the big picture?)

Note: a 2D map is an abstract entity you can invent/create for our own convenience!

Two different uses of a 2D map (both can be combined):

1. **Provide the big picture if analysis outcome can be overlaid.**
2. **Provide a meaningful contextual gateway to a deeper/detailed info.**

Dataset Topic Determined
HIV GSE PrEP
Covid-19/Kaggle
Alzheimer's Disease
Liver Gene Expression
Rain in Australia Kaggle
Water quality
COVID-19 GSE
Twitch.tv
Covid-19/Kaggle
Netflix
U.S. pollution Data
Heart Failure Clinical/Kaggle
DS Job Market/Kaggle
fMRI datasets
Apple Business
Musicnet
Cancer Data
Tweets
Hardware simulator logs
DC Bike Sharing
MAHNOB
MLB Hitting/Pitching
Health statistics

Create Hotspot Links Using Image Maps in the Full Editor

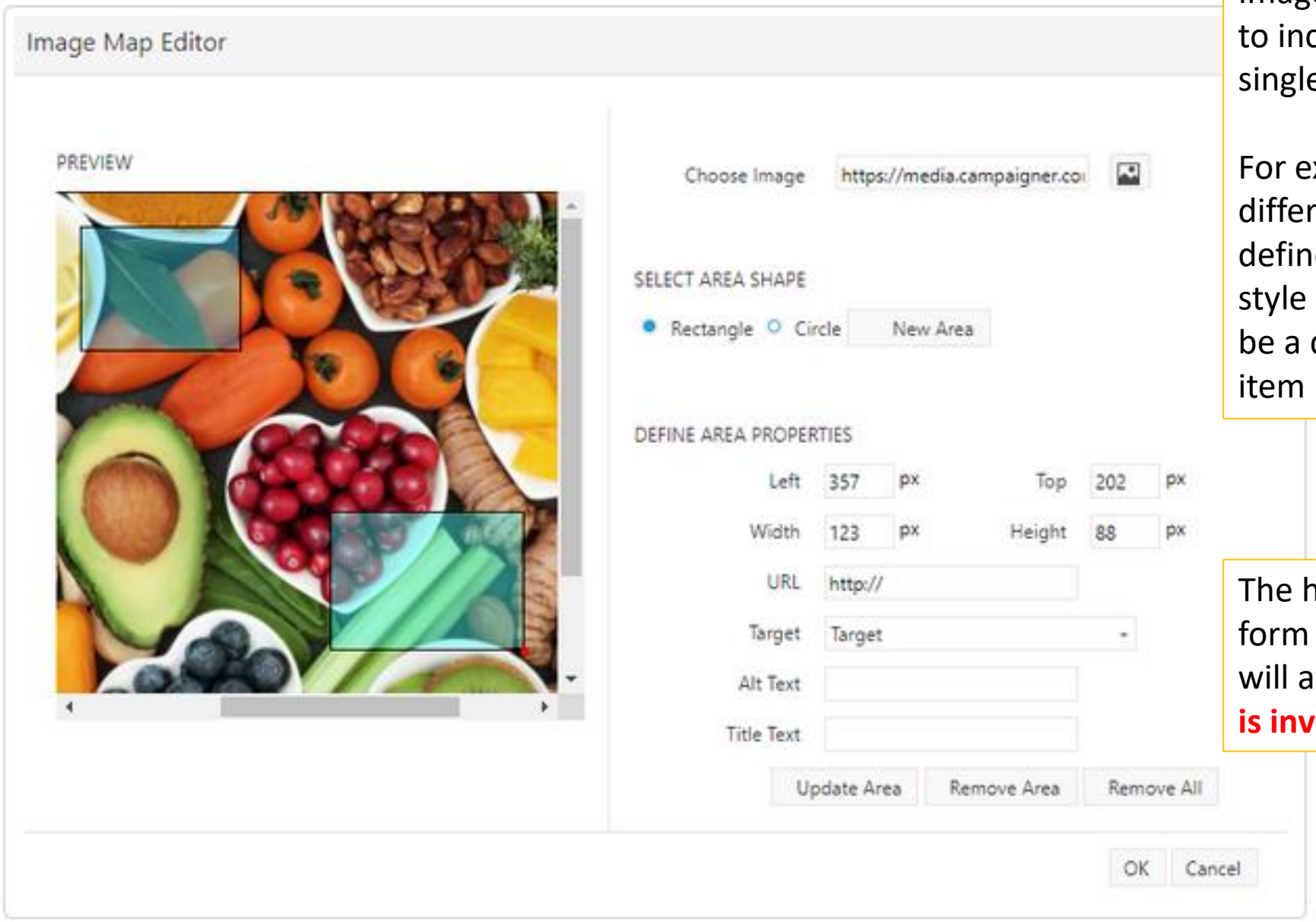


Image mapping provides an easy way to include more than one link on a single image.

For example, if your image displays different styles of clothing, you can define a hotspot on each individual style so that when clicked, there will be a different redirect to that specific item page on your website.

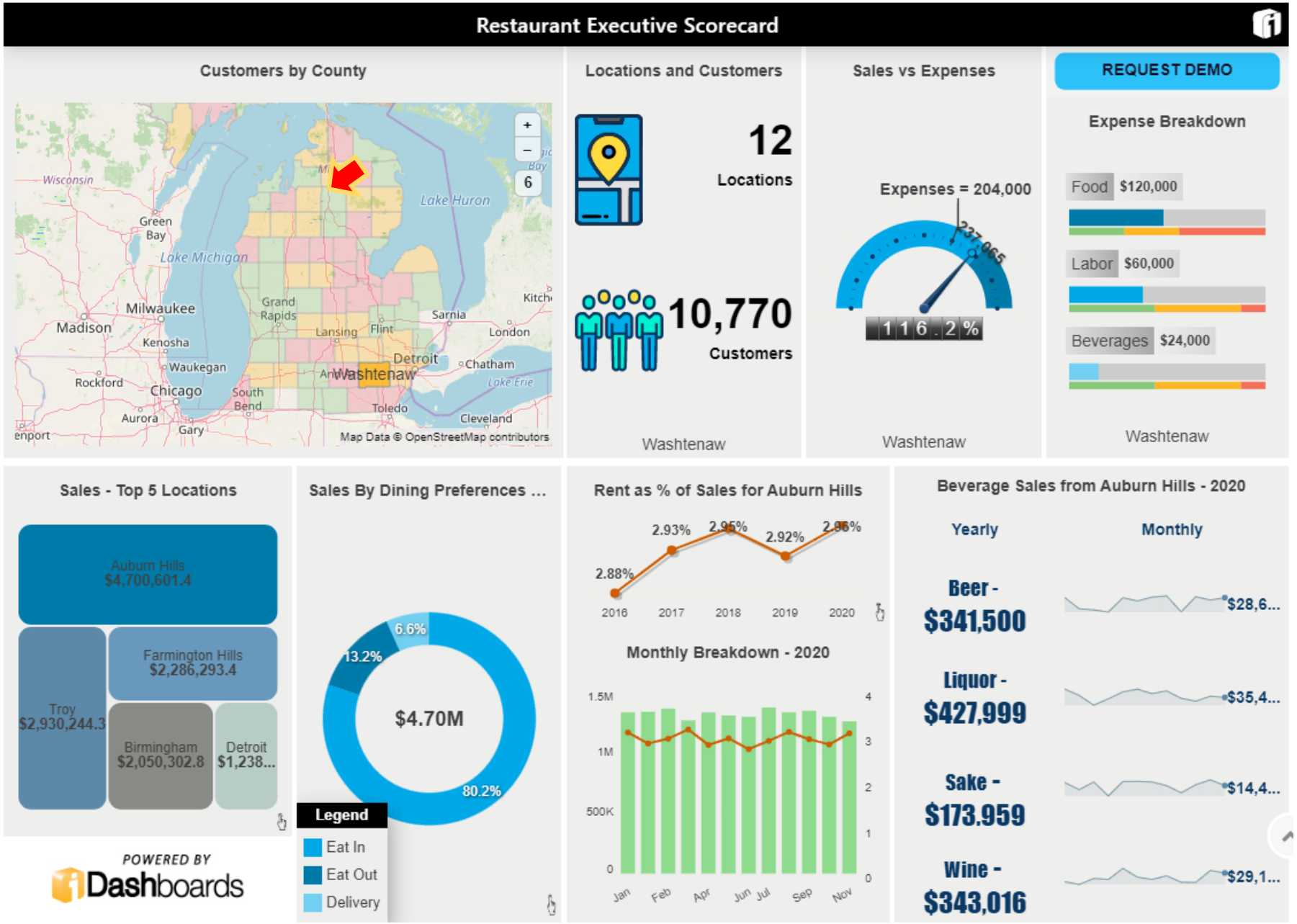
The hotspot is a shaped area in the form of **a rectangle or circle**, that you will add to an image. **The shape itself is invisible to the user.**

Make it visible!!!

https://knowledge.campaigner.com/article/102-create-hotspot-links-using-image-maps

Food & Beverage - Restaurant Executive Scorecard -- Washtenaw County Michigan

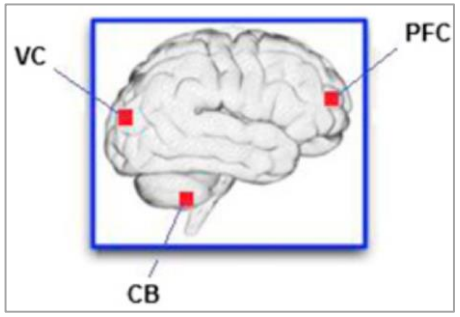
a dynamic SVG map

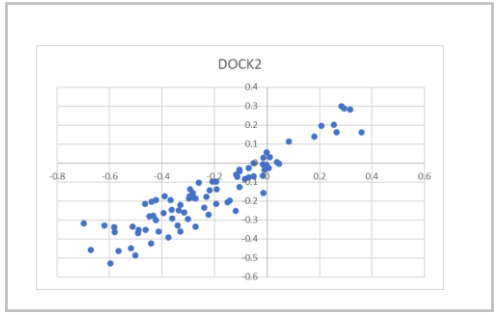


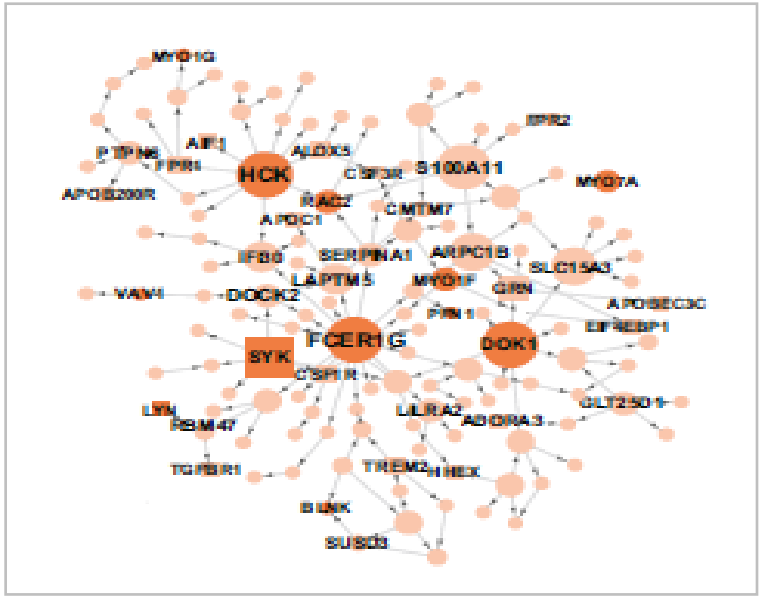
https://www.idashboards.com/dashboard-examples/food-beverage-restaurant-executive-scorecard/

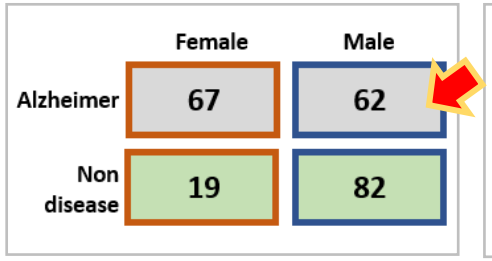
Late-Onset Alzheimer's Disease (LOAD) Dashboard

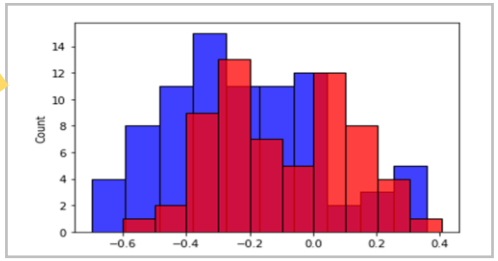
by Sue Smith, CSE5520, Fall 2021

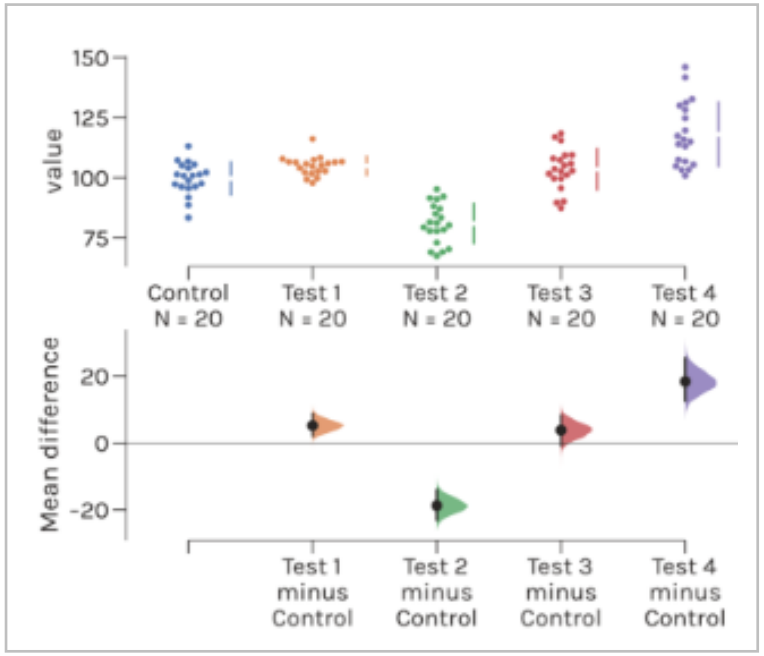


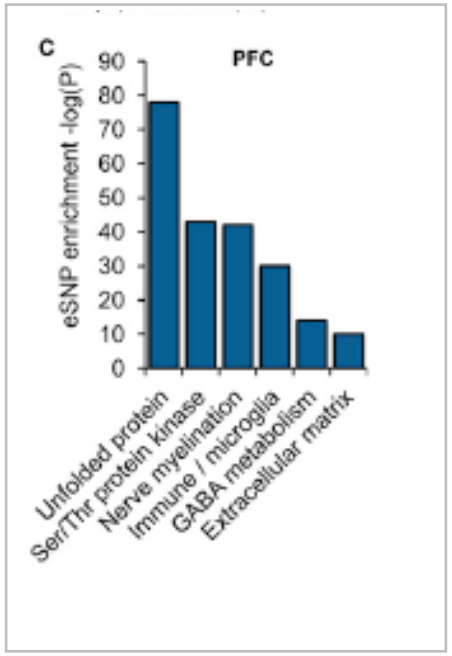


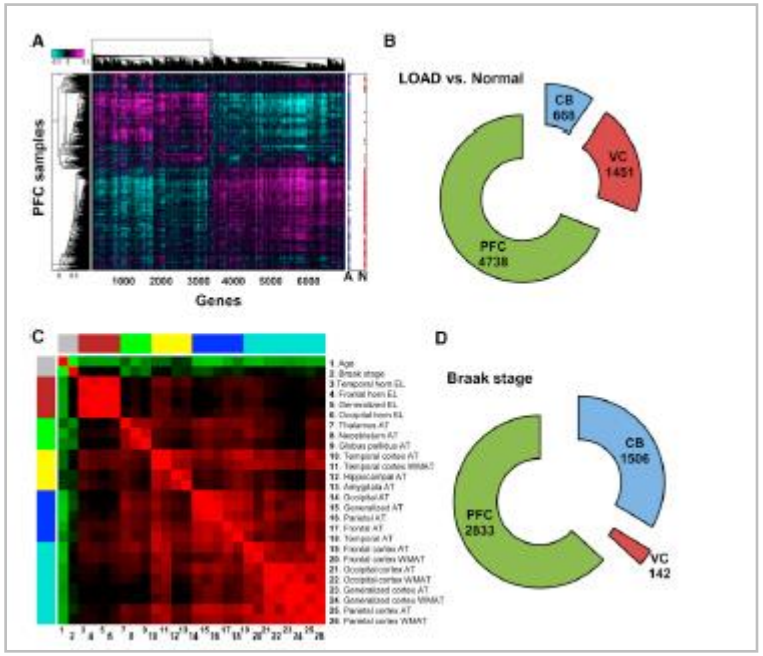












Dashboard
Example

Categorical variable
Discrete variable
Continuous variable

My Project

What data?

What visualization?

What analysis?

Prediction
Causality
Association

Domain: Biology, Engineering, Business, NBA/NFL/MLB, Social Science

Bayesian inference

Basic
Monte Carlo
MCMC

Sampling

Monte Carlo

Classification

Information Gain
KNN
Gaussian mixed model

Clustering

Hierarchical
K-means

Network Analysis

Directional
Non-directional

Kernel Density Estimate

Discrete to continuous

Hypothesis Testing

t- Test
p-value
ROC
FDR

χ^2 Test

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Correlation Analysis

Linear regression
Pearson correlation
Non-linear regression

Scatter plot
Regression line
Correlation matrix
Heatmap

Probability Distribution

Histogram
Boxplot
Violinplot

Bernoulli distribution
Poisson distribution
Gamma distribution
Beta distribution

Data Visualization Road Map

Key to Success in Term Project – Evaluation Criteria

Are you delivering interesting theme(s) that cohesively string together the visualizations that you offer in your Dashboard?

What are you trying to communicate through visualizations you offer from your Dashboard? Is there a message?

How challenging problem(s) are you attempting to solve?

How diversified visualization methods are you delivering? Only one kind?

How creative/unique are your delivered visualization methods?

Are you providing alternative ways of visualizing relationships to contrast strength of each method?

Peer Evaluation form which should be filled and submitted to HuskyCT after both sessions are over.

Private consultation is available during and outside office hrs;
please feel free to contact us via email.