DA 320 Project 7 W. Li

**Part 1**: Review the design tools located at [http://www.creativebloq.com/design-tools/datavisualization-712402?page=1](http://www.creativebloq.com/design-tools/data-visualization-712402?page=1) and [http://www.computerworld.com/article/2507728/enterprise-applications/enterpriseapplications-22-free-tools-for-data-visualization-and-analysis.html](http://www.computerworld.com/article/2507728/enterprise-applications/enterprise-applications-22-free-tools-for-data-visualization-and-analysis.html) Choose a tool, research and play with it, and make a presentation to demo the tool, including a brief introduction, and its advantages and draw backs compare with other tools.

* Make the presentation in **PowerPoint**
* Remember to include your name, class name, and instructor’s name on the first slide, and references on the last slide.
* Make 10 to 15 slides.
* Use ample lists with bullets to make your points clear;
* Include **images sparingly** (**but** **at least** **one**);
* Keep the distracting flash effects (like animation) to a minimum (**at most one**);
* Make it look professional and consider that you could actually use this for a business presentation

**Part 2**: Data is only as good as its completeness. And we have all dealt with incomplete data. But when a dataset has missing or incomplete data, it creates a set of dynamics that are centered on how to handle the missing data. There are, of course, several strategies one can use to address this issue. Some strategies often include:

* Deleting the row where the missing data appears
* Placing a unique number in the missing data field or location
* Averaging the data that DOES appear and placing the average value into the missing data field or location.
* Placing a value in the missing data field or location from an adjacent field or location

Analyze each strategy, and other strategies you may think of, and determine the advantages and disadvantages of each strategy. As part of your analysis, make sure to address if anything can be gained or lost from using one strategy over another.

|  |  |  |
| --- | --- | --- |
| Method | Advantage | Disadvantage |
| Omitting the records or fields with missing values from the analysis | Simplicity  Comparability across analyses | This may be dangerous, as the pattern of missing values may in fact be systematic. Deleting the records with missing values would lead to a biased subset of the data. |
| Replacing the missing value with some constant. | Simplicity | Results in biased estimates and not theoretically driven. |
| Replacing the missing value with the field mean. | Simplicity  Can use complete case analysis methods | The mean may not always be the best choice for what constitutes a representative value. Reduces variability and weakens covariance and correlation estimates in the data. |
| Replacing the missing value with the field mode. | Simplicity  Can use complete case analysis methods | The mode may not always be the best choice for what constitutes a representative value. Reduces variability and weakens covariance and correlation estimates in the data. |
| Replacing the missing values with a value generated at random from the observed distribution of the variable. | A benefit of this method is that the measures of center and spread should remain closer to the original. | Reduces variability and weakens covariance and correlation estimates in the data. |
| Replacing the missing values with imputed values based on the other characteristics of the record.  Regression Imputation | Data and theoretically driven  Uses information from observed data. | Might overestimates correlation estimates and weakens variance. |
| Model-based Methods: Maximum Likelihood  Estimation | Uses full information, both complete cases and incomplete cases, to calculate log likelihood  Unbiased parameter estimates | SEs biased downward |