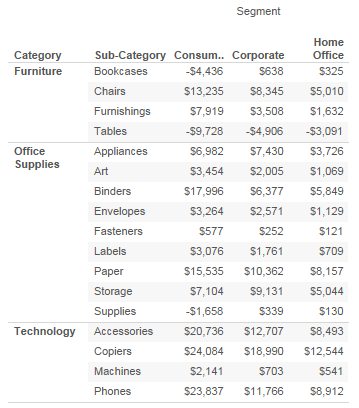
Udacity Tableau Course Notes

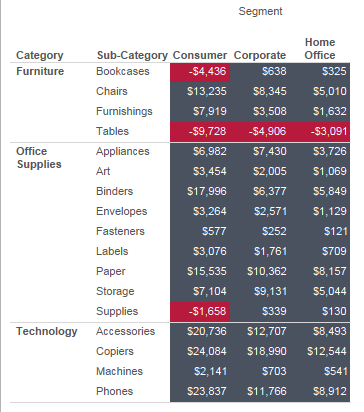
# Fundamentals

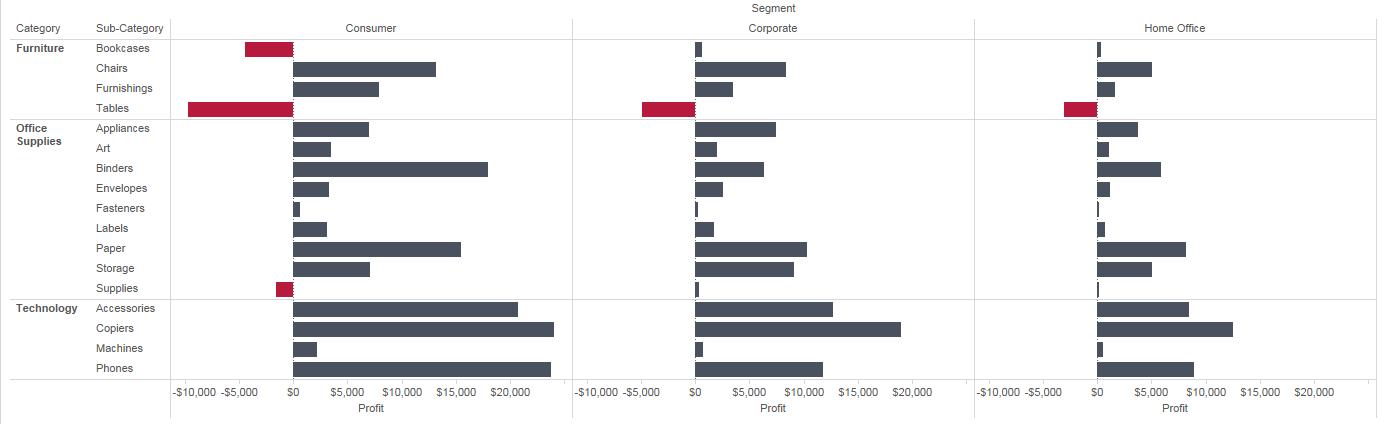
## Data Types

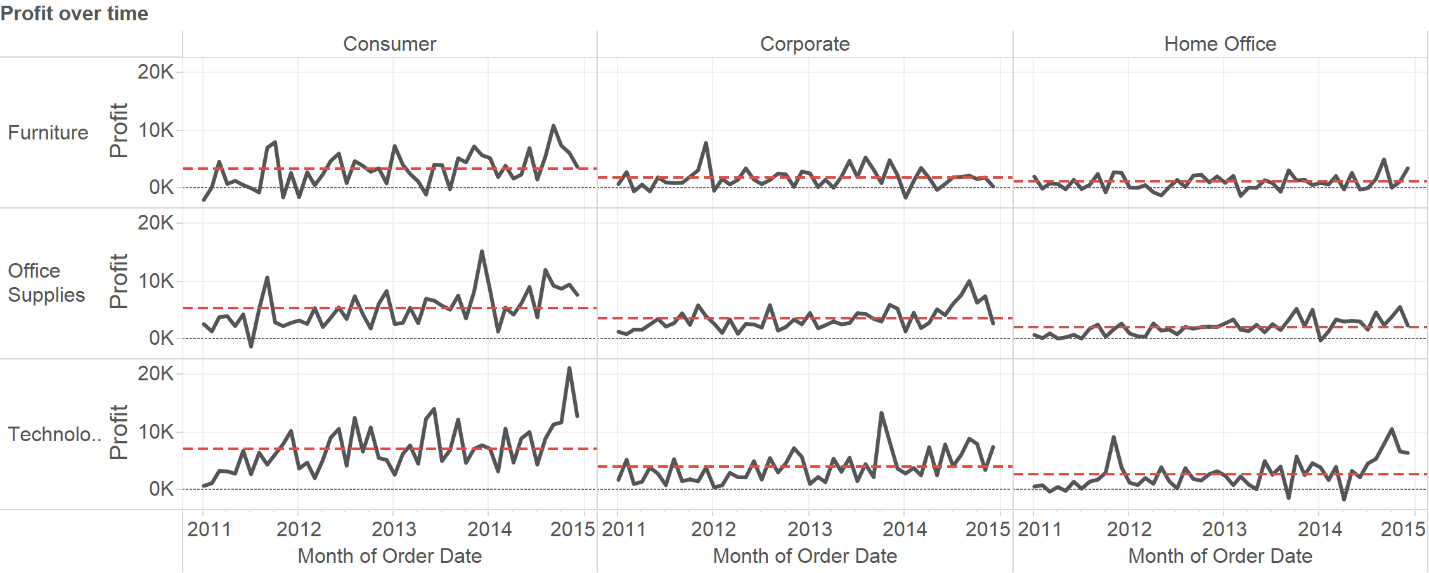
* Quantitative Data
  + **Continuous**: income per person
  + **Discrete**: total population
* Categorical Data
  + **Nominal**: geographical region
  + **Ordered**: population bins, difficulties

## Examples of Data Visualizations









## Visual Encodings

* It is hard to visualize more than 2-D…
* **Retinal variables** can be used to encode additional dimensions

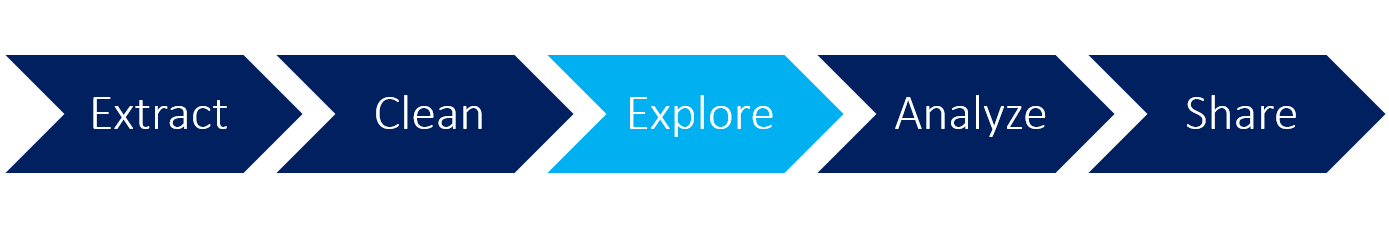
|  |  |
| --- | --- |
| Ordered | Nominal |
| size | color hue |
| orientation | shape |
| color saturation | texture |

* + Time may be animated

## Ranking of Visual Encodings

* From most accurate to less accurate
  + position
  + length
  + angle, slopes
  + area
  + volume
  + color, density
* put the most important variables on axis, choose appropriate encoding is important.

## Exploration vs Explanation



while a lot of time would be spent in dealing with missing data or bad formatted data, the focus of this course would be on exploring datasets.

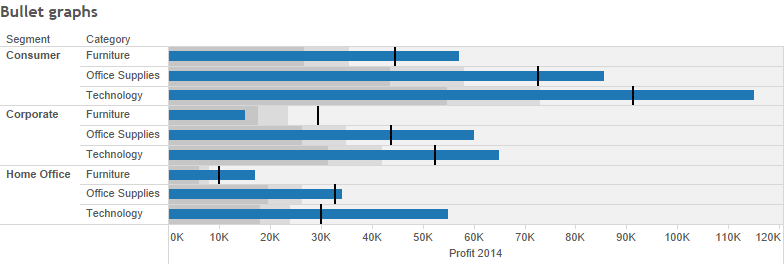
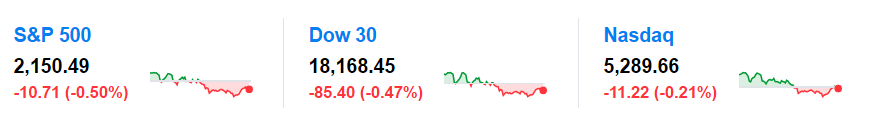
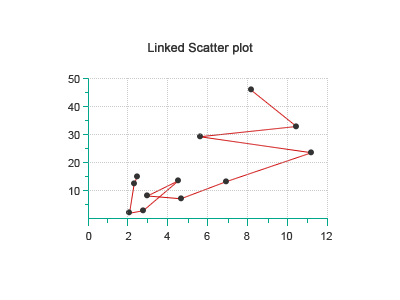
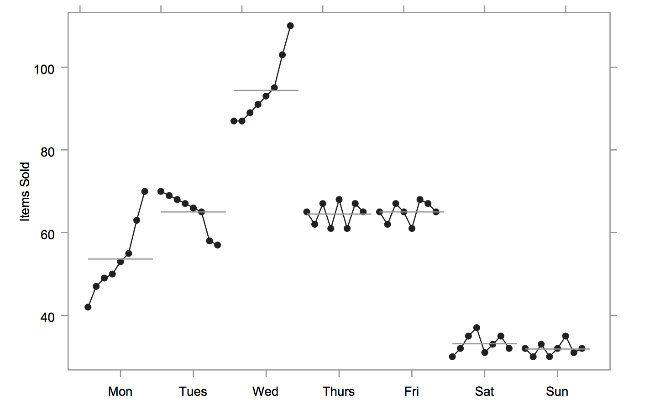
visualization tools often have a easy-to-use versus allowing-more-customization trade off.

\* Data visualizations that tell a story or convey information are considered to be explanatory.

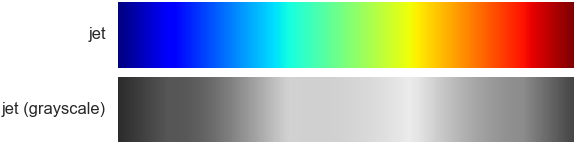
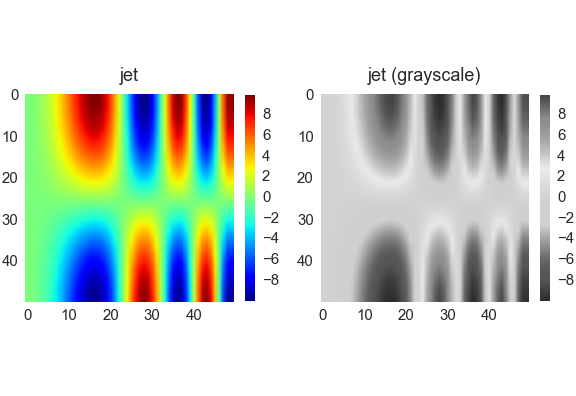
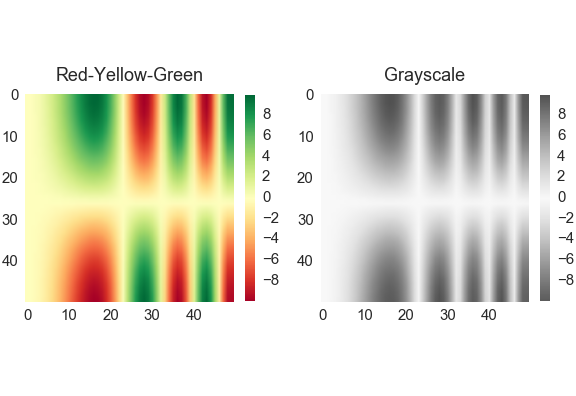
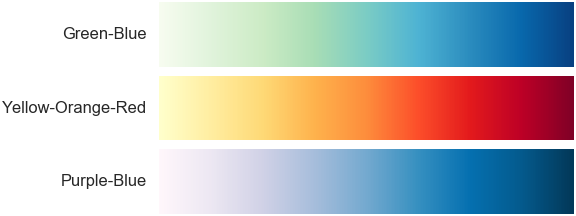
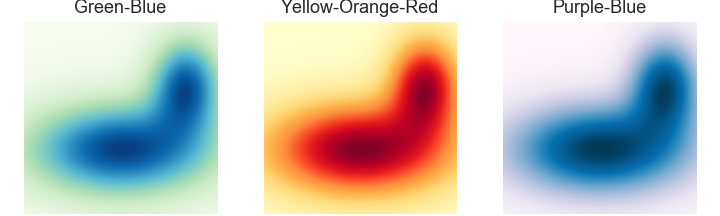
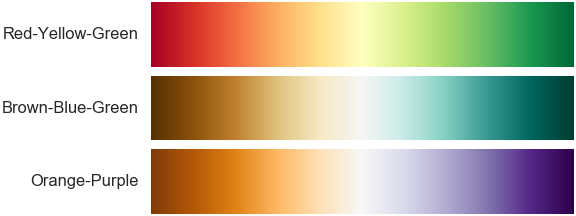
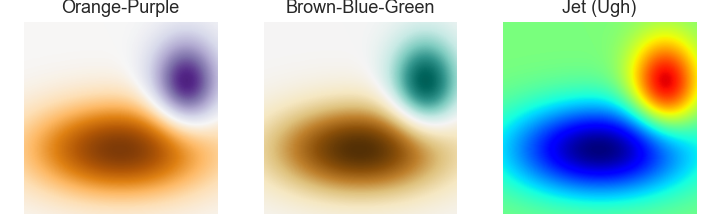
\* Data Visualizations that let you discover trends or patterns in a data set are called exploratory.

# Design Principles

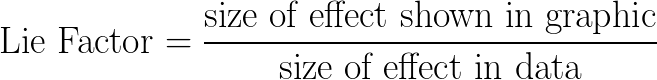
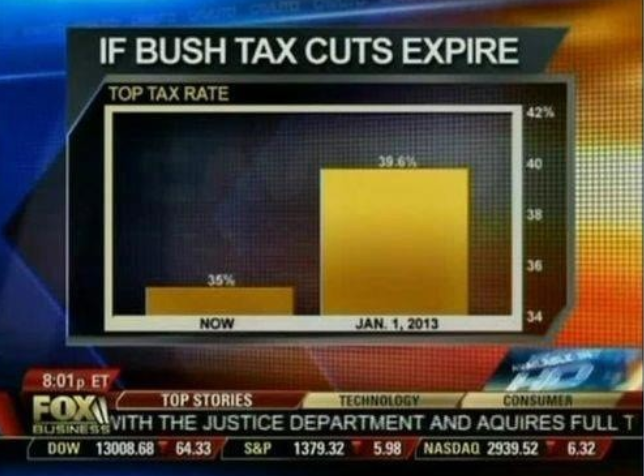
## Chart Types

* Chart types = visual encodings + data types + relationship
* Guide to selecting the right graph: <http://www.perceptualedge.com/articles/ie/the_right_graph.pdf>, <http://www.perceptualedge.com/articles/misc/Graph_Selection_Matrix.pdf>, <http://extremepresentation.typepad.com/.shared/image.html?/photos/uncategorized/choosing_a_good_chart.jpg>, <https://solomonmessing.wordpress.com/2014/10/11/when-to-use-stacked-barcharts/>,
  + There are seven types of quantitative messages
    - **nominal comparison**: bar graph, histogram
    - **time-series**: line charts, connected points show trend
    - **ranking**: ordered bar chart
    - **part-to-whole**: bar chart of percentages, stacked bars can be used to display measure of the whole as well as the parts
    - **deviation**: bars with both positives and negatives
    - **frequency distribution**: histogram, lines may be used to emphasize individual values.
    - **correlation**: scatterplot, bubble chart
* Chart Types:
  + **Bar charts, scatter plots, line plots**: very commonly used
  + **Tables**: can be done with colors!
  + **Geospatial plots**:
    - choropleths: uses color to encode another value associated with the location of the map
    - cartograms: distorts boundaries of regions to encode a value.
  + **Small multiples**: a series of plots with the same scale to make it easier for comparisons between groups.
* Visualizing distributions
  + **Histograms**: for grouping values into ordered groups
  + **Box plots**: common visualization for displaying the general shape of a distribution using intervals.
  + **Violin plots**: display a smooth distribution of the data, approximated using kernel density estimation.
  + **Strip charts**: display actual data for each group as dots, sometimes used on the margins of two-dimensional charts like scatter plots.
* Other awesome graphs
  + **bullet graphs**
    - 
  + **sparkline**
    - 
  + **connected scatterplot**: adds a third dimension such as date to provide extra context
    - 
  + **kernel density estimate**: fancy histogram, smoothed using underlying normal distributions.
  + **cycle plots**: group together data in periodic data
    - 

## Color pallets

* a color pallet is a range of color used to encode data values.
* Jet pallet: The jet/rainbow palette is flawed because the luminance does not transition smoothly from one end to the other. Some colors pop out more than others.
  + 
  + 
* Red-yellow-green: suffers the same problem as jet
  + 
* Sequential palettes: smooth transition from light to dark, or vice versa.
  + 
  + 
  + 
* Diverging palette: suitable for data with some breakpoingt value (negative and positive)
  + 
  + 
* Categories: different color and hue for different categories
* Design for color blindness:
  + red-green blindness: stay away from red-green palettes and try using blue-orange

## Practical Rules

* Use black-white-grey first
* use less intense colors: medium hue, pastel, with higher grey values
* rainbow colors create noisy visualizations that strain the eyes
* use color to highlight
* Stephen Few's Practical Rules for Using Color in Charts
  + Rule #1 - If you want different objects of the same color in a table or graph to look the same, make sure that the background—the color that surrounds them—is consistent.
  + Rule #2 - If you want objects in a table or graph to be easily seen, use a background color that contrasts sufficiently with the object.
  + Rule #3 - Use color only when needed to serve a particular communication goal.
  + Rule #4 - Use different colors only when they correspond to differences of meaning in the data.
  + Rule #5 - Use soft, natural colors to display most information and bright and/or dark colors to highlight information that requires greater attention.
  + Rule #6 - When using color to encode a sequential range of quantitative values, stick with a single hue (or a small set of closely related hues) and vary intensity from pale colors for low values to increasingly darker and brighter colors for high values.
  + Rule #7 - Non-data components of tables and graphs should be displayed just visibly enough to perform their role, but no more so, for excessive salience could cause them to distract attention from the data.
  + Rule #8 - To guarantee that most people who are colorblind can distinguish groups of data that are color coded, avoid using a combination of red and green in the same display.
  + Rule #9 - Avoid using visual effects in graphs.
* The Data-ink ratio = ink used to describe the data / ink used to describe everything else
  + low is bad, high is good
  + low means there are too many extraneous ink.
* The lie-factor
  + 
  + i.e.:
  + 
    - The graphical effect is (146-27)/27 = 4.407, and the data effect is (39.6-35)/35 = 0.131, giving a lie factor of 33.54

# Creating visualizations in Tableau

## About Tableau

* Tableau is a data exploratory tool that allows you to answer some deep questions about your data
* Install: <https://public.tableau.com/s/>
  + the app version is free!

## Notes on Visualizations

* primarily done in tableau file – see folder with Tableau files
* **group**: static list of data points from the all the points
* **set**: dynamic list of data points from all the points, can update with changes to the database.

# Story Telling

## Dashboards

* **Create sheets** to fill a dashboard. Simply drag any related sheets onto a dashboard window.
* **Filters and legends** can be **repositioned** as floating object on top of charts or docked objects to the sides.
* **Creating action filter**: Worksheet->action. Specify source, target and selection clearing behavior of the action.
* Selecting any object within a sheet or dashboard object can bring up **formatting**.
  + Example: tooltip text can be edited to present different colors.
  + The formatting pane can be opened with format menu or right-click dropdown.
    - Font, alignment, shading, borders, and line are common formatting objects.
    - formatting can be applied at the sheet scope, or rows and/or column scope.

## Stories

* Simply drag the sheets to each story tab to form a story.
* Edit each story tab to narrate a story
* Think about an argument/hypothesis
  + Then check if it the data support or reject the hypothesis
  + Present the data
* An example: sales in third world country is going down, hypothesis is prices are too low, check the prices after discount, too much discount, did discount program work to increase revenue? yes -> try to lower discount to lift profit, no -> try to get rid of discount program all together as it is not working, alternatively -> cut cost for those stores.