



Stony Brook University

Visualization of Student Migration Data Using Google Charts Sankey Diagrams

Sean V. Hoffman

Institutional Research Analyst

Office of Institutional Research, Planning & Effectiveness

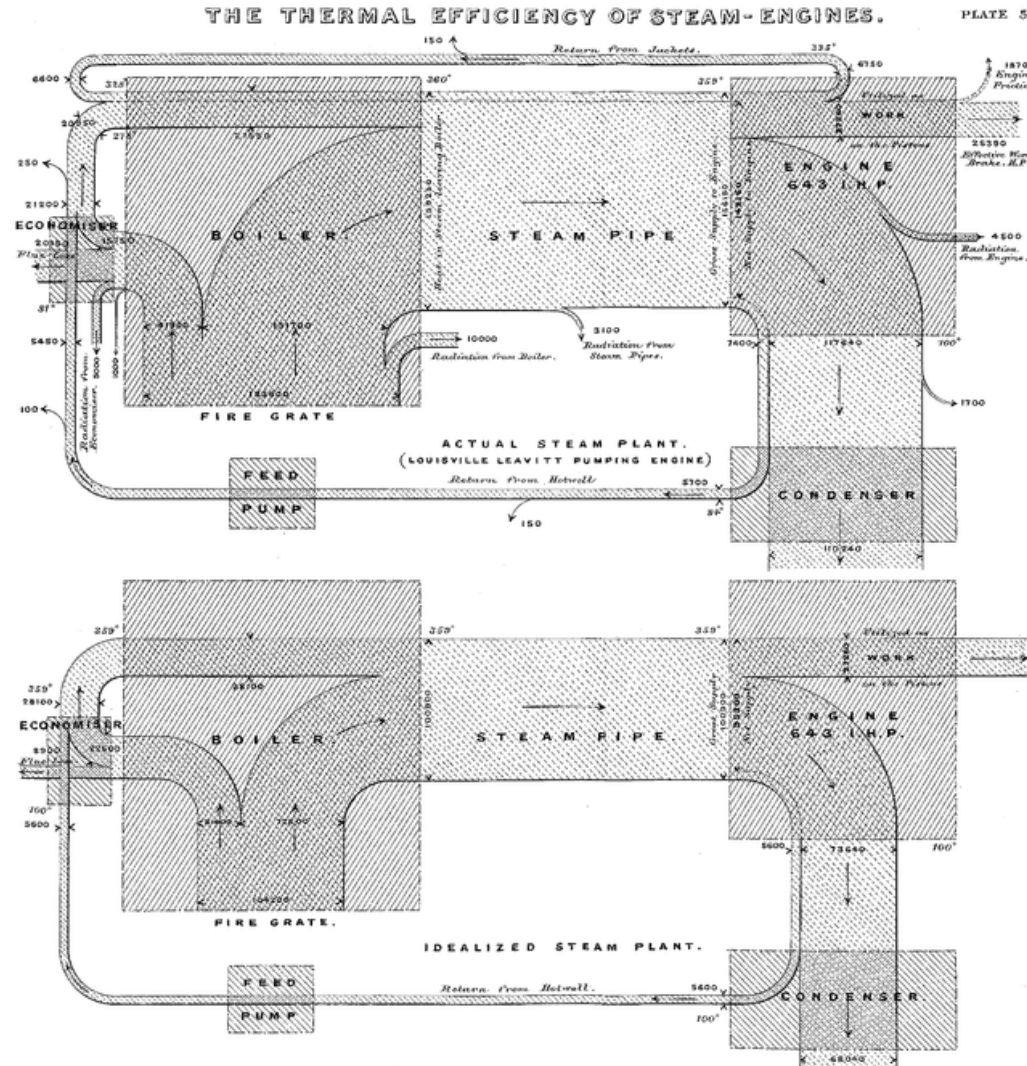
- Goals
 - To learn the basics of Sankey diagrams
 - To gain an understanding of how Sankey diagrams can be used in Institutional Research
 - To be able to create a basic Sankey diagram in Google Charts, using Excel data
 - To review an example of a Sankey diagram created to visualize student migration at Stony Brook

- Introduction to Sankey diagrams
- Sankey Diagrams and Institutional Research
- Building an Example Sankey
 - Discuss the process behind collecting data and creating a basic Sankey diagram
 - Look further into how a more complex diagram is created
- Google Charts for Institutional Research
 - Other charts available
 - Dashboards in Google Charts

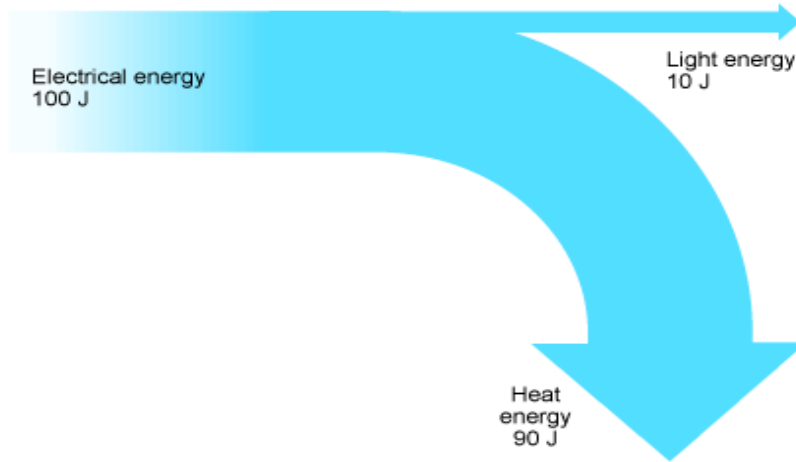


WHAT IS A SANKEY DIAGRAM?

- Originally created to show efficiency of a steam engine by Captain Matthew Henry Phineas Riall Sankey in 1898 (Source: Wikipedia)
- Flow diagram
- Used to visualize flow of a system, transfer of items, movement of goods...
- Thickness of line represents magnitude of flow

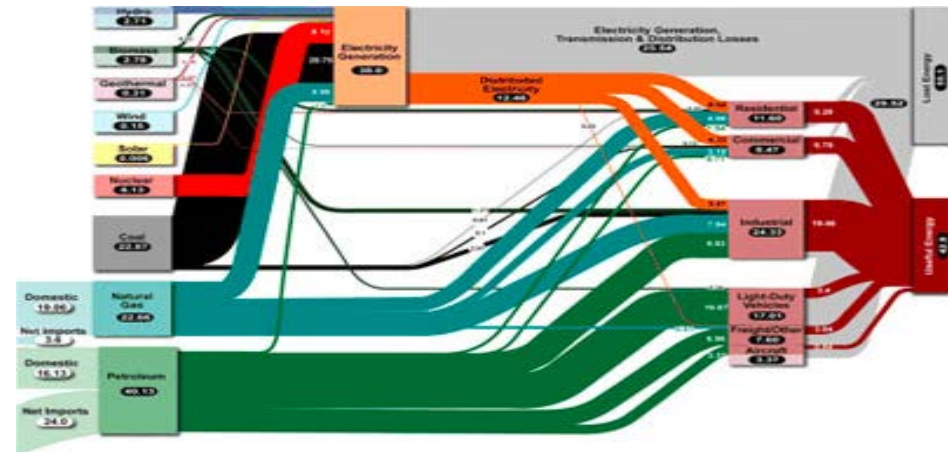


EXAMPLES OF SANKEY DIAGRAMS



Simple Sankey Diagram of a Lightbulb

Multi Level Sankey Energy Diagram

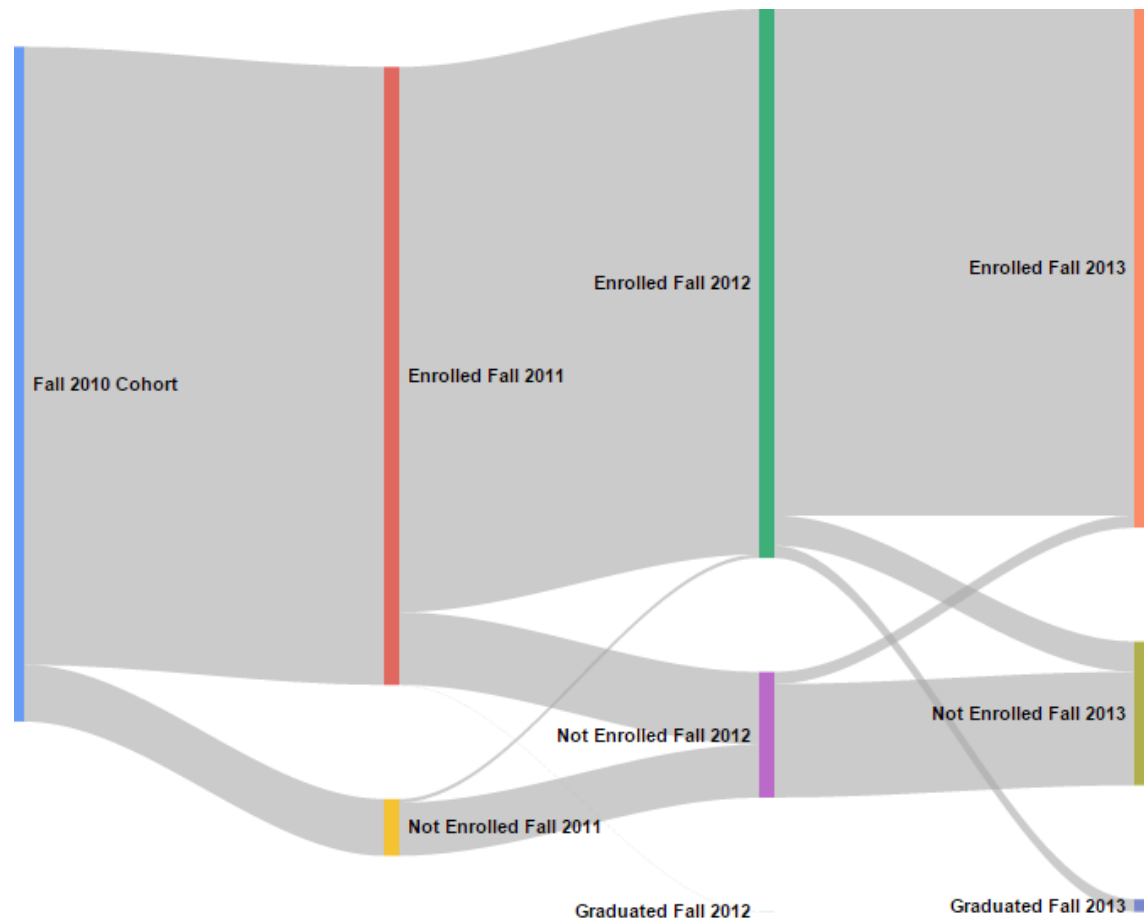




- **Nodes:** Represent sources and destinations of flows
 - Cities, Universities, Academic Departments, Accounts
- **Links:** Links indicate that there exists flow between the nodes it connects
- **Weight:** Also known as the thickness. The higher the weight of the link, the thicker it is

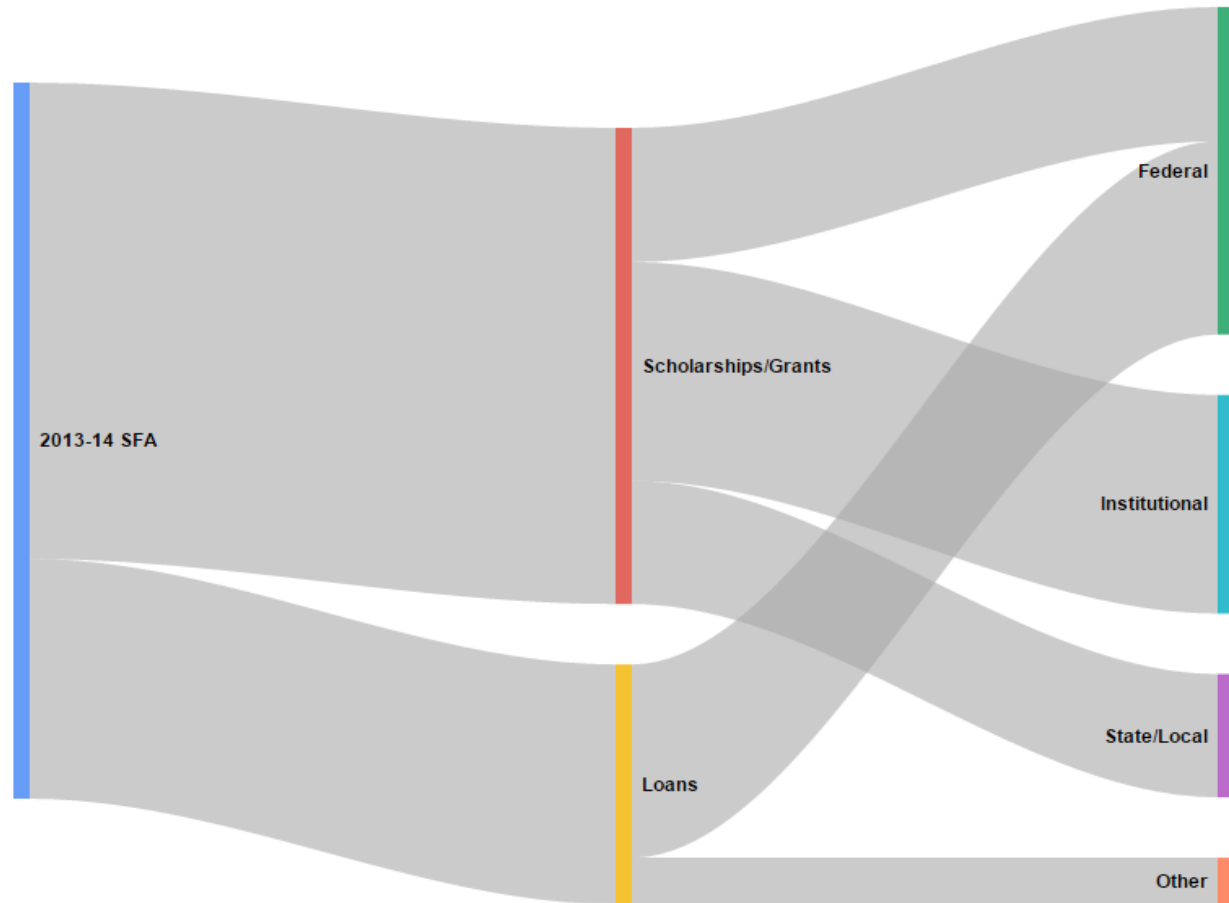


- Useful where there is movement
- Student transitions
 - Major changes, Time to degree, Transfer destinations
- Transactional data
 - Finance, Student Financial Aid, Research Expenditures, HR Data
- Example: Tracking a cohort (Fall 2010 SBU FTFT to their 4th Fall)

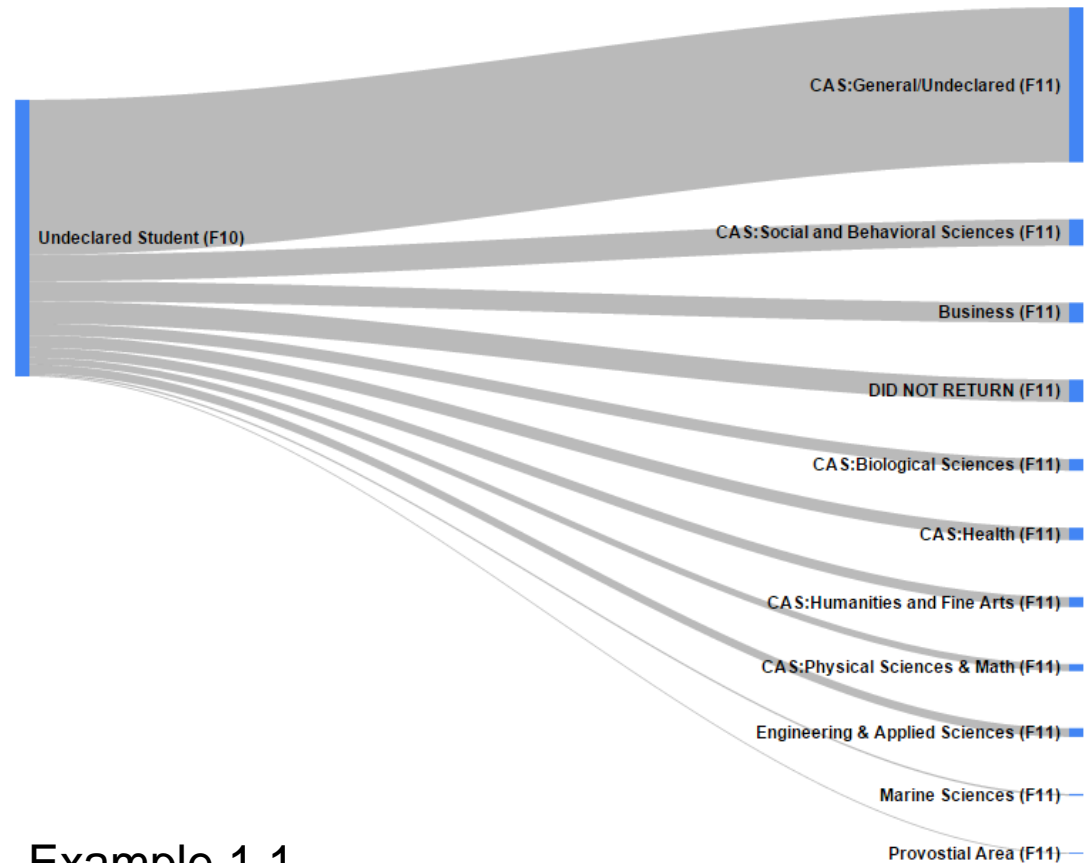


Example 1

- 2013 – 14 IPEDS SFA Data from Stony Brook
- Total financial aid
- Scholarships/Grants vs Loans
- Aid source

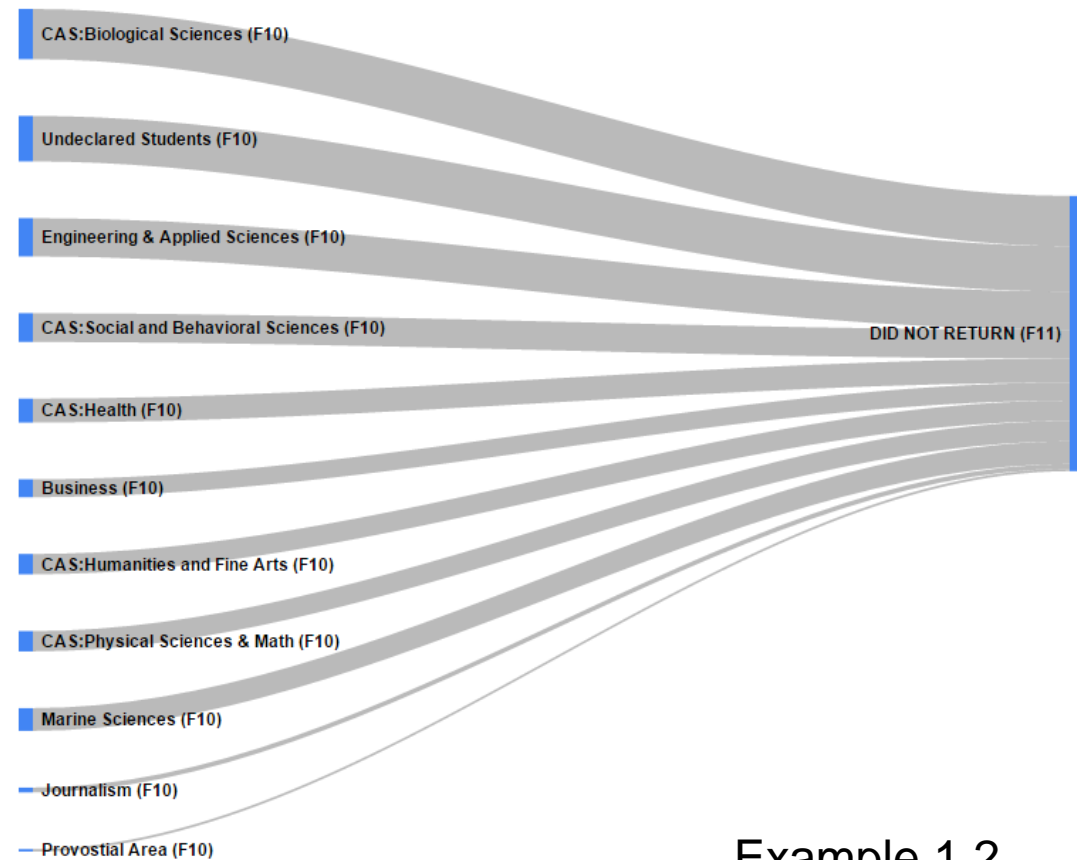


- Closer look at Stony Brook University FTFT 2010 students
 - Looking by Major groupings
- Looking for retention/graduation interventions: undeclared?
- For now, look at one year change (Fall to Fall)
- Immediate observations
 - Majority do not leave
- Questions about those not returning:
 - In line with other majors?



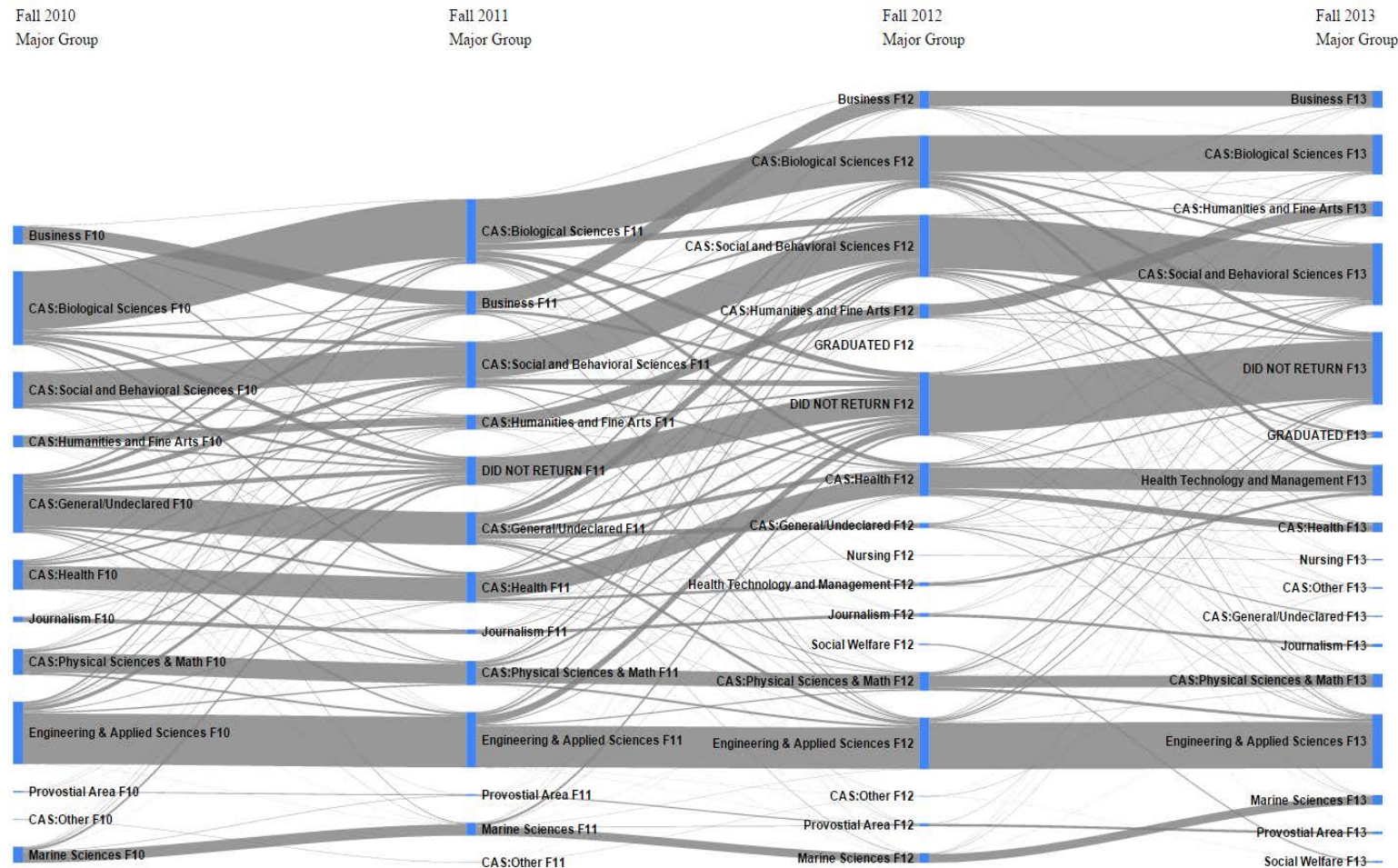
Example 1.1

- To answer this question, let's look at students who don't return
- Observations
 - Not out of line with other majors
- Benefits
 - Not a list of numbers and percentages
 - Easy to visualize



Example 1.2

Migration of Stony Brook University Fall 2010 FTFT Students by Major Group



- Add major groupings to initial cohort tracking example (Example 1)

Example 1.3



- What does this diagram show us? What does it show decision makers?
 - Helps to show complexity to decision makers and those making requests
 - Demonstrate trends in flow, or “currents”
 - Highlight problem areas: high transfer out, high attrition,...
 - Can also show successes
 - Prompt further investigation into previously unnoticed subgroups or trends
- SBU case specific examples
 - SOMAS – High out, low in
 - CEAS – AMS in, engineering out
 - Undeclared students persist, do not leave at as significantly different rates as thought
 - Major alone will not give us a target

- Several options for creating Sankey diagrams (some others listed at end)
- Sankey diagram included as part of Google Charts
 - Clean, straightforward, and does not require a background in coding (**I am not a programmer!**)
 - Replicable and easily editable to accommodate data changes or new project
 - Can be combined to create dashboards with other Google Charts
- <https://developers.google.com/chart/interactive/docs/gallery/sankey>
- We will examine the code behind some simple Sankey diagrams
 - Creating the diagram only involves changing **one** section of the code provided by Google
 - Users can modify more if they like



```
<html>
<head>
  <script type="text/javascript" src="https://www.google.com/jsapi"></script>
  <script type="text/javascript">
    google.load("visualization", "1.1", {packages:["sankey"]});
    google.setOnLoadCallback(drawChart);

    function drawChart() {
      var data = new google.visualization.DataTable();
      data.addColumn('string', 'From');
      data.addColumn('string', 'To');
      data.addColumn('number', 'Weight');
      data.addRows([
        ['Business (F10)', 'DID NOT RETURN (F11)', 15],
        ['CAS:Biological Sciences (F10)', 'DID NOT RETURN (F11)', 42],
        ['Undeclared Students (F10)', 'DID NOT RETURN (F11)', 38],
        ['CAS:Health (F10)', 'DID NOT RETURN (F11)', 20],
        ['CAS:Humanities and Fine Arts (F10)', 'DID NOT RETURN (F11)', 17],
        ['CAS:Physical Sciences & Math (F10)', 'DID NOT RETURN (F11)', 17],
        ['CAS:Social and Behavioral Sciences (F10)', 'DID NOT RETURN (F11)', 24],
        ['Engineering & Applied Sciences (F10)', 'DID NOT RETURN (F11)', 32],
        ['Journalism (F10)', 'DID NOT RETURN (F11)', 4],
        ['Marine Sciences (F10)', 'DID NOT RETURN (F11)', 19],
        ['Provostial Area (F10)', 'DID NOT RETURN (F11)', 2]
      ]);
      var options = {
        height: 600,
        width: 750,
        sankey: { node: { nodePadding: 40, label: {fontName: 'Arial',
          fontSize: 12, bold: true, color: 'black'} } }
      };
      var chart = new google.visualization.Sankey(document.getElementById('sankey_basic'));
      chart.draw(data, options);
    }
  </script>
</head>
<body>
  <div id="sankey_basic" style="width: 900px; height: 900px;"></div>
</body>
</html>
```

```
<html>
<head>
  <script type="text/javascript" src="https://www.google.com/jsapi"></script>
  <script type="text/javascript">
    google.load("visualization", "1.1", {packages:["sankey"]});
    google.setOnLoadCallback(drawChart);

    function drawChart() {
      var data = new google.visualization.DataTable();
      data.addColumn('string', 'From');
      data.addColumn('string', 'To');
      data.addColumn('number', 'Weight');
      data.addRows([
        ['Business (F10)', 'DID NOT RETURN (F11)', 15],
        ['CAS:Biological Sciences (F10)', 'DID NOT RETURN (F11)', 42],
        ['Undeclared Students (F10)', 'DID NOT RETURN (F11)', 38],
        ['CAS:Health (F10)', 'DID NOT RETURN (F11)', 20],
        ['CAS:Humanities and Fine Arts (F10)', 'DID NOT RETURN (F11)', 17],
        ['CAS:Physical Sciences & Math (F10)', 'DID NOT RETURN (F11)', 17],
        ['CAS:Social and Behavioral Sciences (F10)', 'DID NOT RETURN (F11)', 24],
        ['Engineering & Applied Sciences (F10)', 'DID NOT RETURN (F11)', 32],
        ['Journalism (F10)', 'DID NOT RETURN (F11)', 4],
        ['Marine Sciences (F10)', 'DID NOT RETURN (F11)', 19],
        ['Provostial Area (F10)', 'DID NOT RETURN (F11)', 2]
      ]);
      var options = {
        height: 600,
        width: 750,
        sankey: { node: { nodePadding: 40, label: {fontName: 'Arial',
          fontSize: 12, bold: true, color: 'black'} } }
      };
      var chart = new google.visualization.Sankey(document.getElementById('sankey_basic'));
      chart.draw(data, options);
    }
  </script>
</head>
<body>
  <div id="sankey_basic" style="width: 900px; height: 900px;"></div>
</body>
</html>
```

This is the only section that needs to change: **The links and weights**

Some example code can be found on Google's webpage

- Basic Steps
 1. Define nodes, links, and weights
 2. Gather data (enrollments for this example)
 3. Create code for nodes and links using data
 4. Insert the code, replacing old nodes and links
 5. Run code in HTML editor, you're done!
- Going Further
 - Edit Google chart options
 - Edit HTML options (*not covered in this presentation*)
- Let's try it

- Build the Major Migration Sankey from Example 1.2
- The following structure is used
 - Nodes: Major groups by fall terms (Ex: Engineering Fall 2010)
 - Links: Links represent movement between the two major groups they connect (the year of the movement is also specified by which nodes are connected – we limit to one year movements)
 - Weights: Denotes the amount of students moving between the two program groups in the represented time frame
- Not all definitions will be the same. Ensure you will be able to demonstrate what you are trying to show
 - Initial attempts at this example led to adding a year component to the node labeling
 - Initial cohort node – helps with start distribution

- Need enrollment data, stored in a spreadsheet
 - Pulled from enrollment records
- Create a panel that shows enrollment in one fall and then subsequent fall enrollment
 - For SBU, data stored on major, school/college, graduation status, college/division (local grouping of majors to reduce number of links)
- Summary sheet for total changes: will use to create code
 - Pivot table

- “Rows” create nodes, links, and their relationship

- Columns inside each row are:

- Source – Where does the flow start?
- Destination – Where does the flow go?
- Value – How much flow?

Data Format

Rows: Each row in the table represents a connection between two labels. The third column indicates the strength of that connection, and will be reflected in the width of the path between the labels.

Columns:

	Column 0	Column 1	Column 2
Purpose:	Source	Destination	Value
Data Type:	string	string	number
Role:	domain	domain	data

- Row code format: ['Source', 'Destination', Value],

- ['Source', 'Destination', Value], *Note: ORDER MATTERS*

Must start and end with open bracket: [

Include single quotes around node names: ' '

Separate source, destination and value with commas. Also, separate **EACH** row entry with a comma (except for the final entry): ,

- Use a concatenate formula in your spreadsheet to create rows in this format

```
data.addRow([
  ['Business (F10)', 'DID NOT RETURN (F11)', 15],
  ['CAS:Biological Sciences (F10)', 'DID NOT RETURN (F11)', 42],
  ['Undeclared Students (F10)', 'DID NOT RETURN (F11)', 38],
  ['CAS:Health (F10)', 'DID NOT RETURN (F11)', 20],
  ['CAS:Humanities and Fine Arts (F10)', 'DID NOT RETURN (F11)', 17],
  ['CAS:Physical Sciences & Math (F10)', 'DID NOT RETURN (F11)', 17],
  ['CAS:Social and Behavioral Sciences (F10)', 'DID NOT RETURN (F11)', 24],
  ['Engineering & Applied Sciences (F10)', 'DID NOT RETURN (F11)', 32],
  ['Journalism (F10)', 'DID NOT RETURN (F11)', 4],
  ['Marine Sciences (F10)', 'DID NOT RETURN (F11)', 19],
  ['Provostial Area (F10)', 'DID NOT RETURN (F11)', 2]
]);
```

Leave →

Replace ←

Leave →

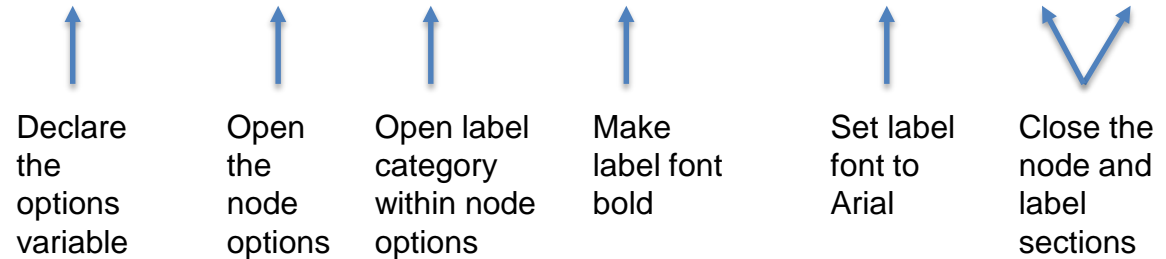
This sample is from example 1.2. You can use any previous Sankey code to start.

- Copy basic example code from Google Sankey page or code from a previous diagram you created
- Place in editor (Notepad, Notepad++, HTML online editor, JSFiddle...)
- Leave the first and last lines
 - data.addRow([
 -]);
- Remove links in between these lines, and add your links
 - Make sure last link does **NOT** end with a comma, the code will not work

- Notepad
 - Save as .htm, open file with internet browser
- Notepad++
 - Save file, use run option to run in a browser
- HTML editor
 - <http://htmledit.squarefree.com/> - updates in real time
- JS Fiddle
 - <https://jsfiddle.net/> - need some HTML experience, since Javascript and HTML is separated by this editor

- Options are entered with the following format

- Example: `var options = {node: {label: {bold: true, font: 'Arial'} ...} ...}`



- Available option categories

- Sankey (Node, Link, and Iterations subcategories – Node and Link are shown on next page)
- Height
- Width
- `forceFrame`
- Tooltips (can be used to change hover effect)



```
sankey: {
  link: {
    color: {
      fill: '#efd', // Color of the link.
      fillOpacity: 0.8, // Transparency of the link.
      stroke: 'black' // Color of the link border.
      strokeWidth: 1 // Thickness of the link border (default 0).
    }
  }
}
```

```
sankey: {
  node: {
    label: {
      fontName: 'Times-Roman',
      fontSize: 12,
      color: '#000',
      bold: true,
      italic: false
    },
    labelPadding: 6, // Horizontal distance between the label and the node.
    nodePadding: 10, // Vertical distance between nodes.
    width: 5 // Thickness of the node.
  }
}
```


- Sankey
 - Link
 - Color
 - Fill
 - FillOpacity
 - Stroke
 - StrokeWidth
- Sankey
 - Node
 - Label
 - FontName
 - FontSize
 - Color
 - Bold
 - Italic
 - LabelPadding
 - NodePadding
 - Width

- Multi level Sankey Diagrams are coded in the same manner
 - Keep adding links
 - Logical flow (Example, F10 to F11, F11 to F12,...)
 - Add in order for organization, Google will add them in best fit
- SBU migration has hundreds of links
 - Still created in the same manner
 - Pivot tables and concatenate in Excel
- Options will be key for Visualization
 - Use node padding and sizing, label padding and sizing
 - Change chart size

OTHER CHARTS AVAILABLE FROM GOOGLE CHARTS

Traditional Graphs	Diagrams
Area Charts (Traditional and Stepped)	Bubble Charts
Bar Charts	Box and Whisker Plots (Candlestick Charts)
Column Charts	Calendar Charts
Combo Charts	Gauge Charts
Histograms	Geographic Charts
Intervals	Organizational Charts
Line Charts	Tables
Pie Charts	Timelines
Scatter Charts	Tree Map Charts
Time Series (Annotated)	Word Trees
Trend lines	**User created community charts are also available**

- Dashboards allow combination of charts and controls
- Controls act as filters
 - Category (Pick from available), String (Search), ChartRange, DateRange, NumberRange
 - Customizable filter options (Examples: starting states, allow one choice only...)
- Can use same data source across multiple charts and filters, or multiple data sources
- Can control one or many charts with filter; can use multiple filters per chart



Stony Brook University

Search SB

Institutional Research Planning & Effectiveness

HOME

Home

Undergraduate Student Degree Outcomes and Major Migration Fall-to-Fall

These charts depict the degree outcomes and major migration of undergraduate students from the census date of a fall semester to the census date of the following fall semester. The charts in part answer the question "where are these students one year later?" Pull-down menus allow for selection of a given year, a specific major/department, and a specific student academic level: U0 (Non-matriculated), U1 (Freshman) 0-23 credits earned, U2 (Sophomore) 24-56 credits earned, U3 (Junior) 57-84 credits earned, U4 (Senior) 85+ credits earned. The bottom chart shows a university comparison for a given year and academic level. It is important to observe that various academic and departmental policies, the distribution of student across academic levels, and time frames will affect these figures and major-to-major comparisons are not valid in many cases.

Department:

*All Departments ▾

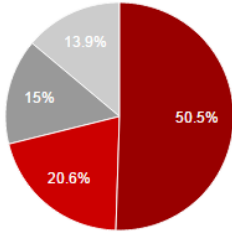
Year:

*All Years ▾

Level:

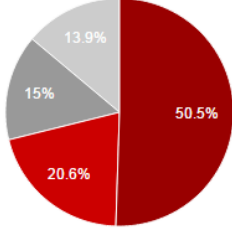
*All Levels ▾

Fall to Fall Undergraduate Retention, Graduation, Migration, and Exit by Department, Year, and Level




Outcome	Percentage
Remained in Department	50.5%
Graduated from Stony Brook	20.6%
Left Department, Still Enrolled	15%
Did not Enroll at Stony Brook	13.9%

Fall to Fall Undergraduate Retention, Graduation, Migration, and Exit by Level



Outcome	Percentage
Remained in Department	50.5%
Graduated from Stony Brook	20.6%
Left Department, Still Enrolled	15%
Did not Enroll at Stony Brook	13.9%


Stony Brook University
Institutional Research Planning & Effectiveness

HOME

Home

Undergraduate Student Degree Outcomes and Major Migration Fall-to-Fall

These charts depict the degree outcomes and major migration of undergraduate students from the census date of a fall semester to the census date of the following fall semester. The charts in part answer the question "where are these students one year later?" Pull-down menus allow for selection of a given year, a specific major/department, and a specific student academic level: U0 (Non-matriculated), U1 (Freshman) 0-23 credits earned, U2 (Sophomore) 24-56 credits earned, U3 (Junior) 57-84 credits earned, U4 (Senior) 85+ credits earned. The bottom chart shows a university comparison for a given year and academic level. It is important to observe that various academic and departmental policies, the distribution of student across academic levels, and time frames will affect these figures and major-to-major comparisons are not valid in many cases.

Department:

Applied Mathematics and Statistics

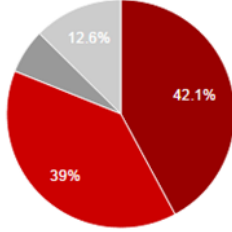
Year:

2013 to 2014

Level:

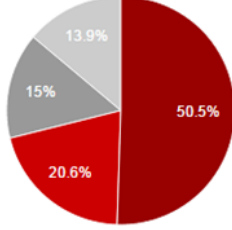
*All Levels

Fall to Fall Undergraduate Retention, Graduation, Migration, and Exit by Department, Year, and Level



Category	Percentage
Remained in Department	42.1%
Graduated from Stony Brook	39%
Left Department, Still Enrolled	12.6%
Did not Enroll at Stony Brook	8.3%

Fall to Fall Undergraduate Retention, Graduation, Migration, and Exit by Level



Category	Percentage
Remained in Department	50.5%
Graduated from Stony Brook	20.6%
Left Department, Still Enrolled	15%
Did not Enroll at Stony Brook	13.9%

Department Category Filter set to Applied Math

Year Category Filter set 2013-14

Second pie chart set to only change with "Level" Category Filter

Connecting to Data

- Google charts allows connection to your database with php and a .json file
- https://developers.google.com/chart/interactive/docs/php_example

Other Resources for Creating Sankey Diagrams

- D3.js
 - Javascript Visualization library
 - Powerful, with more options than Google Charts
 - More complex
 - <http://d3js.org/>
- Tableau
 - Requires manipulation
 - Not flexible, very complicated



- Thank you!
- Examples will be posted on our Webpage within the next week
 - <http://www.stonybrook.edu/commcms/irpe/dashboards/viz.html>
- Contact information
 - sean.hoffman@stonybrook.edu