# Who am I?

- Mark Davenport – Sr Research Associate, IR, UNCG
- I started computer programming in 1982; BASIC and COBOL
- My programming career lasted 6 months
- This fact will become relevant by the end of the presentation

# Why am I here?

- Analytics and visualization are becoming increasingly important in higher ed
  - Patterns of major change from application through graduation are one important example

- IR must learn to do more with less
  - For every costly software solution, there is likely an open-source (free) alternative

# Why major change?

- Can be an indicator of program viability
- Can provide predictive metrics for monitoring growth/decline, changing need for resources
- Can provide insights into changes in student interests/needs
- Can provide insights into retention/graduation anomalies

# Monitoring program viability/change

- Do the students that start in a program stay with that program?
- How long does it take them to graduate?
- Do programs have unknown 'feeder' programs?
- Which programs loose the most students between Fall 1 and Fall 2?  Fall 2 and Fall 3?

# Change in student interest

- If students leave the program, where do they go (e.g., pre-professional majors that don't make the cut)?

- Is interest in particular majors dwindling/ballooning?

- Stop-out/drop-out patterns?

# Outcome anomalies: An example

- Geography program at UNCG
- Fewer than 5 start in Geography as FT-FY new students but 20 of 30 may graduate
- Grad rate based only on the FT-FY cohort may swing wildly due to low numbers
- Geography picks up many 2nd- and 3rd-year from other data-intensive majors like computer science
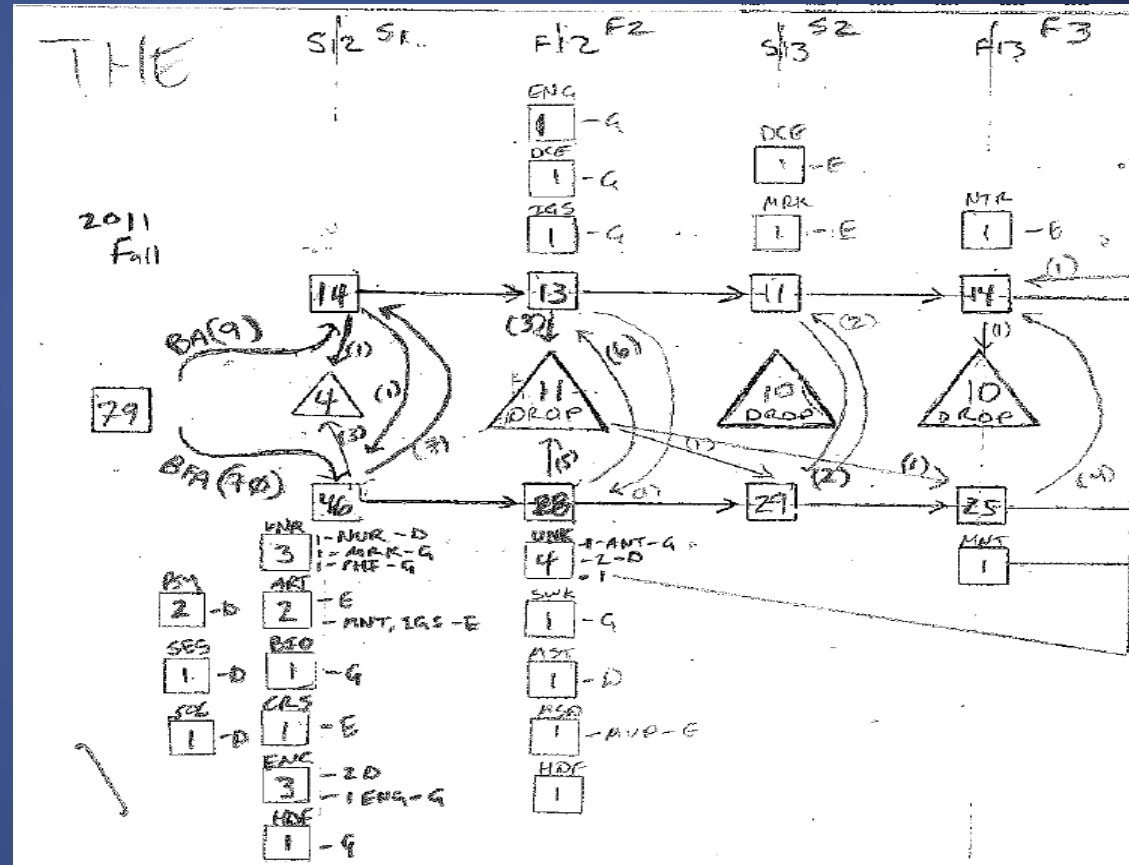
# An early attempt to map major change

## The raw data

| | E | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|
| | | | DEPTF1_CNT | | CONCF1_CNT | | DEGF1_CNT | CONCDEGF1_CNT |
| 1 | GRADTERM | CONCDEG_STRING | | CONC_F1 | | DEG_F1 | | |
| 2 | 201501 | U104BAU104BAU104BAU104BAU104BAU104BAU104BAGGGGGGGG | 84 | U104 | 5 | BA | 26 | 5 |
| 3 | 201501 | U104BAU104BAU104BAU104BAU155BAU155BAU155BAU155BAGGGGGGGG | 84 | U104 | 5 | BA | 26 | 5 |
| 4 | 201508 | U104BAU104BAU189BAU190BAU190BAU190BAU190BAGGGG | 84 | U104 | 5 | BA | 26 | 5 |
| 5 | | U104BAU105BADDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD | 84 | U104 | 5 | BA | 26 | 5 |
| 6 | 201508 | U104BAU331BSU331BSU331BSU331BSU331BSU331BSU331BSU331BSGGGG | 84 | U104 | 5 | BA | 26 | 5 |
| 7 | | U105BADDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD | 84 | U105 | 21 | BA | 26 | 21 |
| 8 | | U105BADDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD | 84 | U105 | 21 | BA | 26 | 21 |
| 9 | | U105BADDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD | 84 | U105 | 21 | BA | 26 | 21 |

# An early attempt to map major change

## The visualization

- Fall 2011 – 79 FT-FY new Theater students in 2 programs

- Boxes contain major headcounts in a fall term; triangles are stop/drop headcounts

- Arrows show flow between box/triangle elements

# An early attempt to map major change

# The problem?

- Tedious to create
- Impossible to explain

# So, how do you visualize change

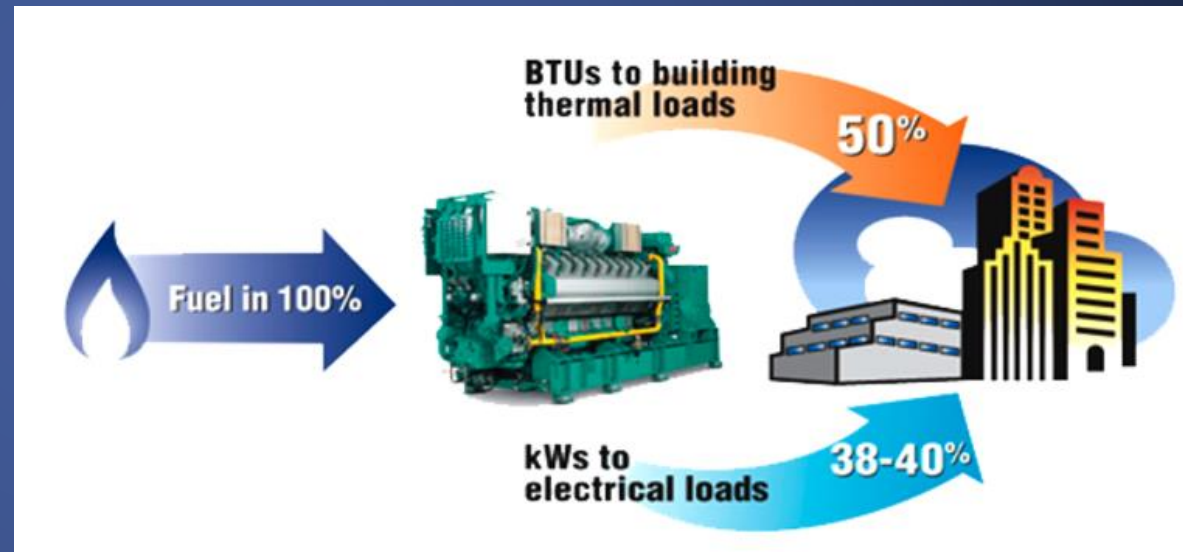## Visualization must show change as an aspect of resource flow through a system



Image source: http://aselca.com.co/en/environment/

# So, how do you visualize change

This subway chart shows many groups (as colored subway lines), but traffic might have been better represented making the lines larger/smaller between stops rather than inserting dots of various size.
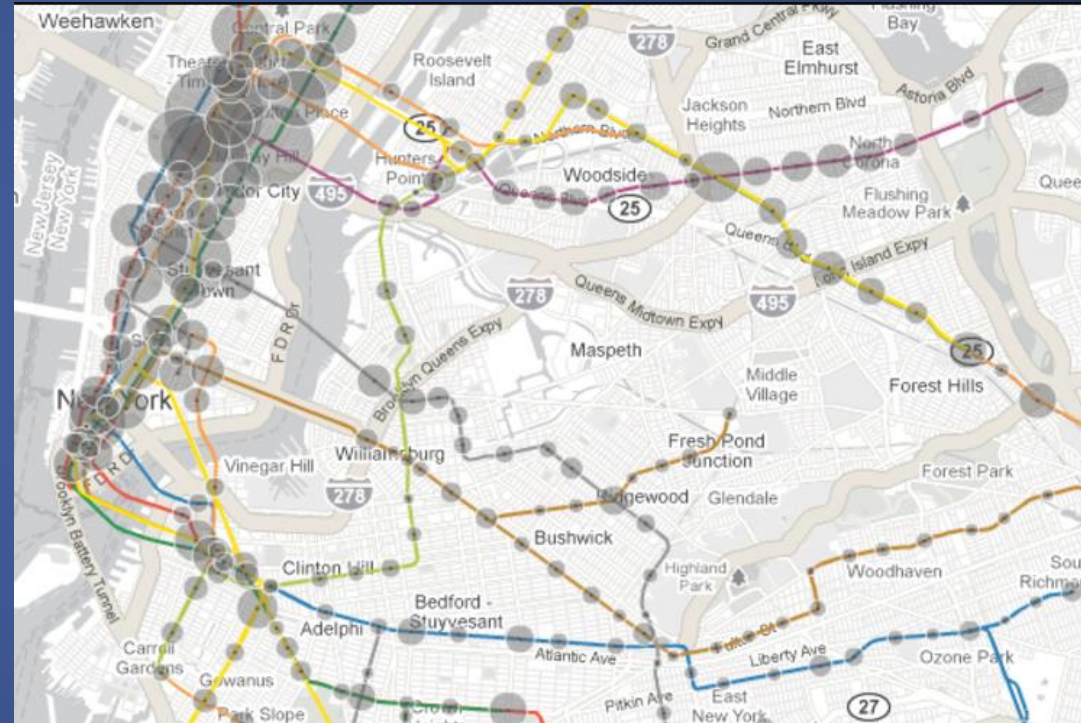


Image source: http://research.prattsils.org/blog/coursework/information-visualization/visualizing-mta-subway-ridership-on-subway-map/

# So, how do you visualize major change

## Our requirements:

- Linear time dimension

- Many, many groups at once

- Groups change in size from one time period to another

- Viz must document movement from group-to-group

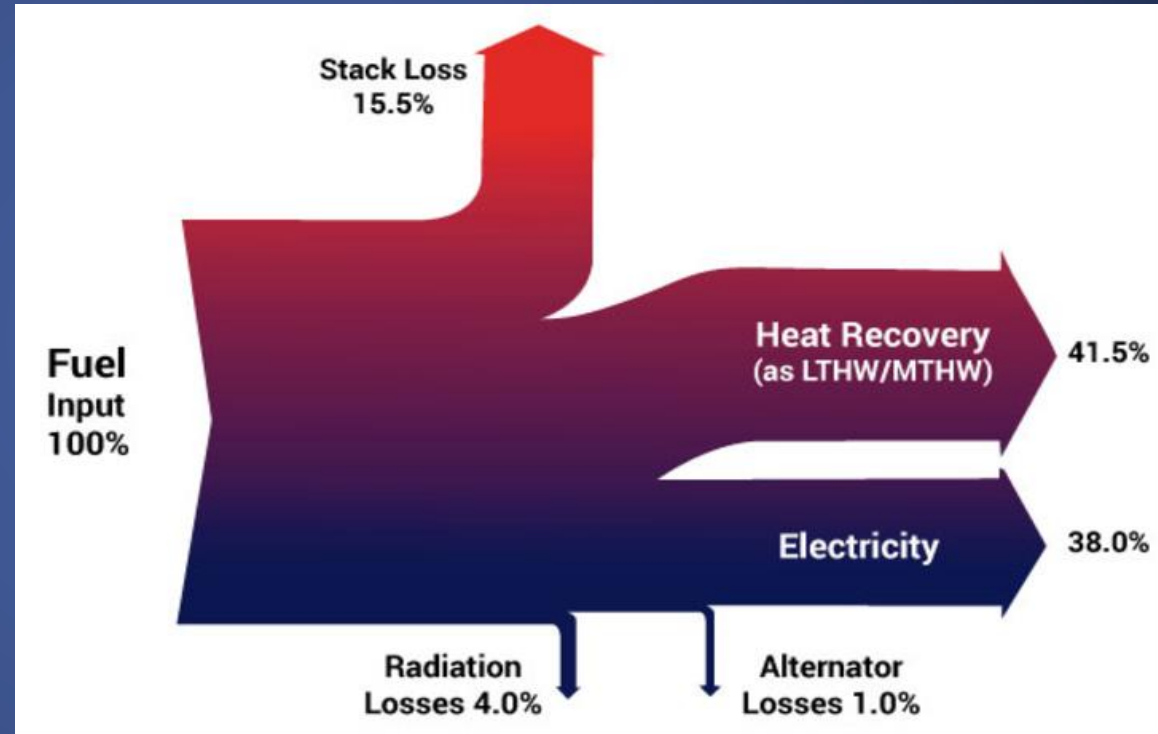- Ideally, viz can be scaled to show school/department-level changes



Image source: http://aselca.com.co/en/environment/

# M-O-A-D-F-V*:  The Sankey Diagram

Takes its name from Matthew Sankey and his diagram demonstrating the efficiency of a steam engine

The defining characteristics:

1)  greater flow = wider path
2)   lesser flow = narrower path
3)  Paths indicate direction of flow

* Mother of all dynamic flow visualizations



THE THERMAL EFFICIENCY OF STEAM-ENGINES.
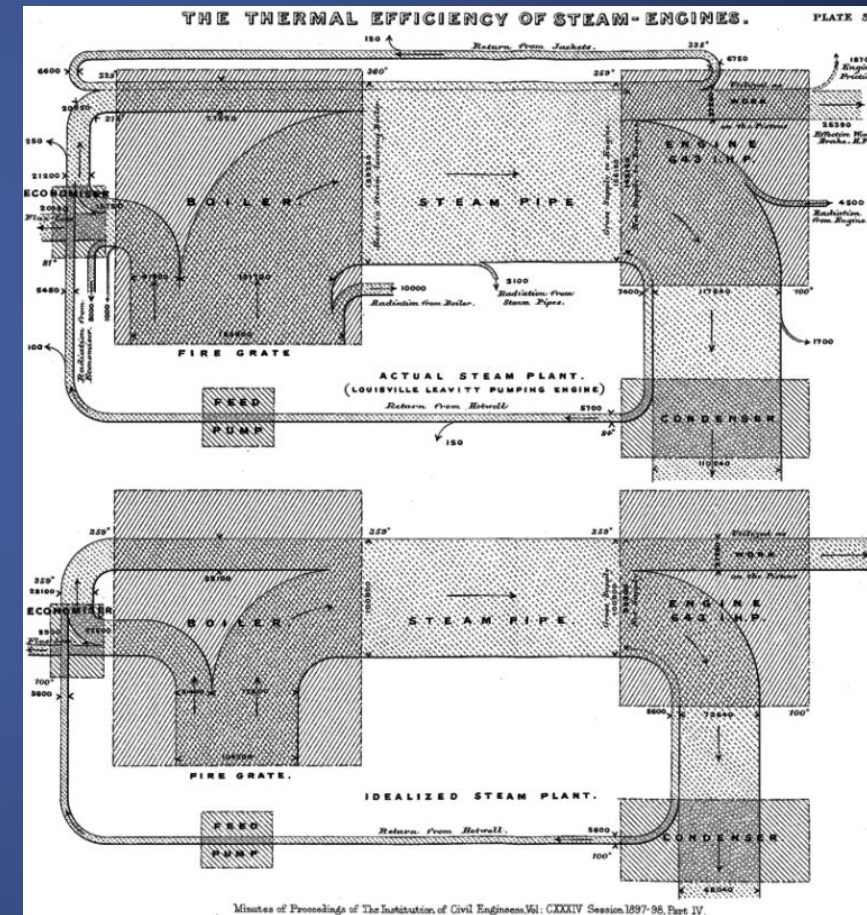
# The Prototypic Sankey Diagram

Charles Joseph Minard (1781-1870) French civil engineer and early pioneer of data visualization created this graphic to describe Napoleon's troop losses in Russia in 1812-13.
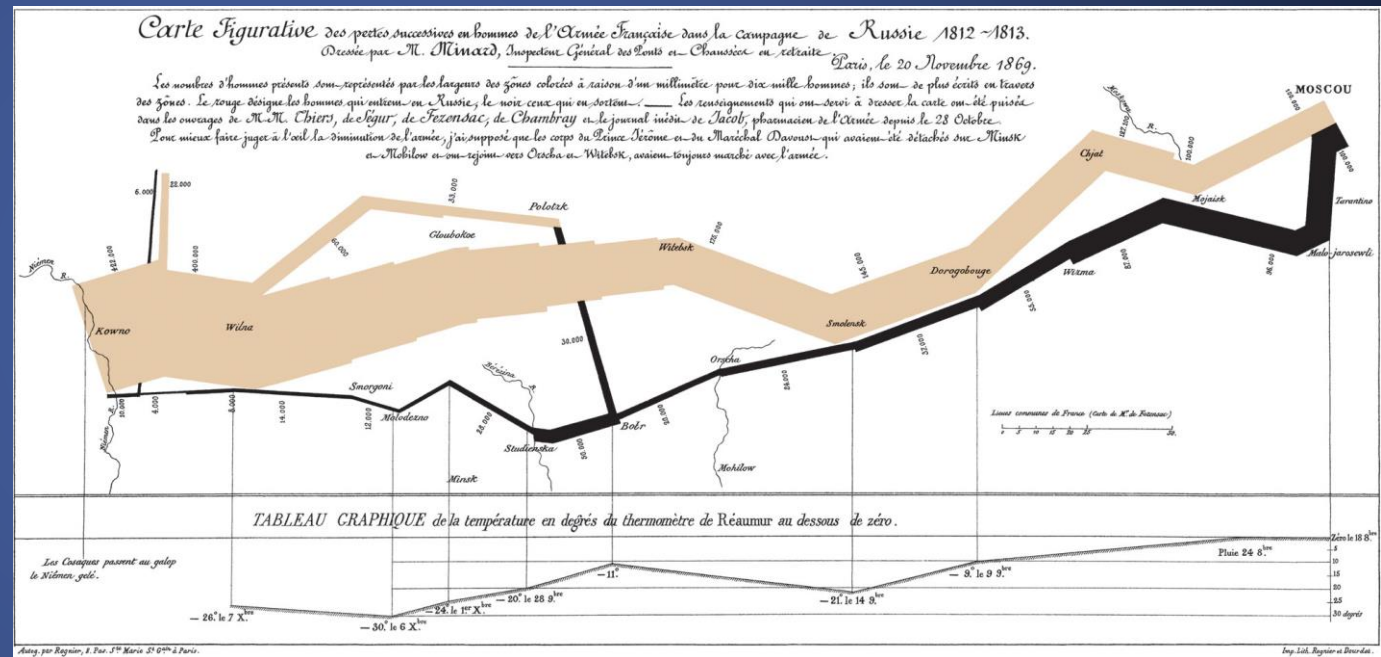


Image source: https://en.wikipedia.org/wiki/Charles_Joseph_Minard

The chart body represents a conceptual map of the route of the march from the Neiman river to Moscow, and back.
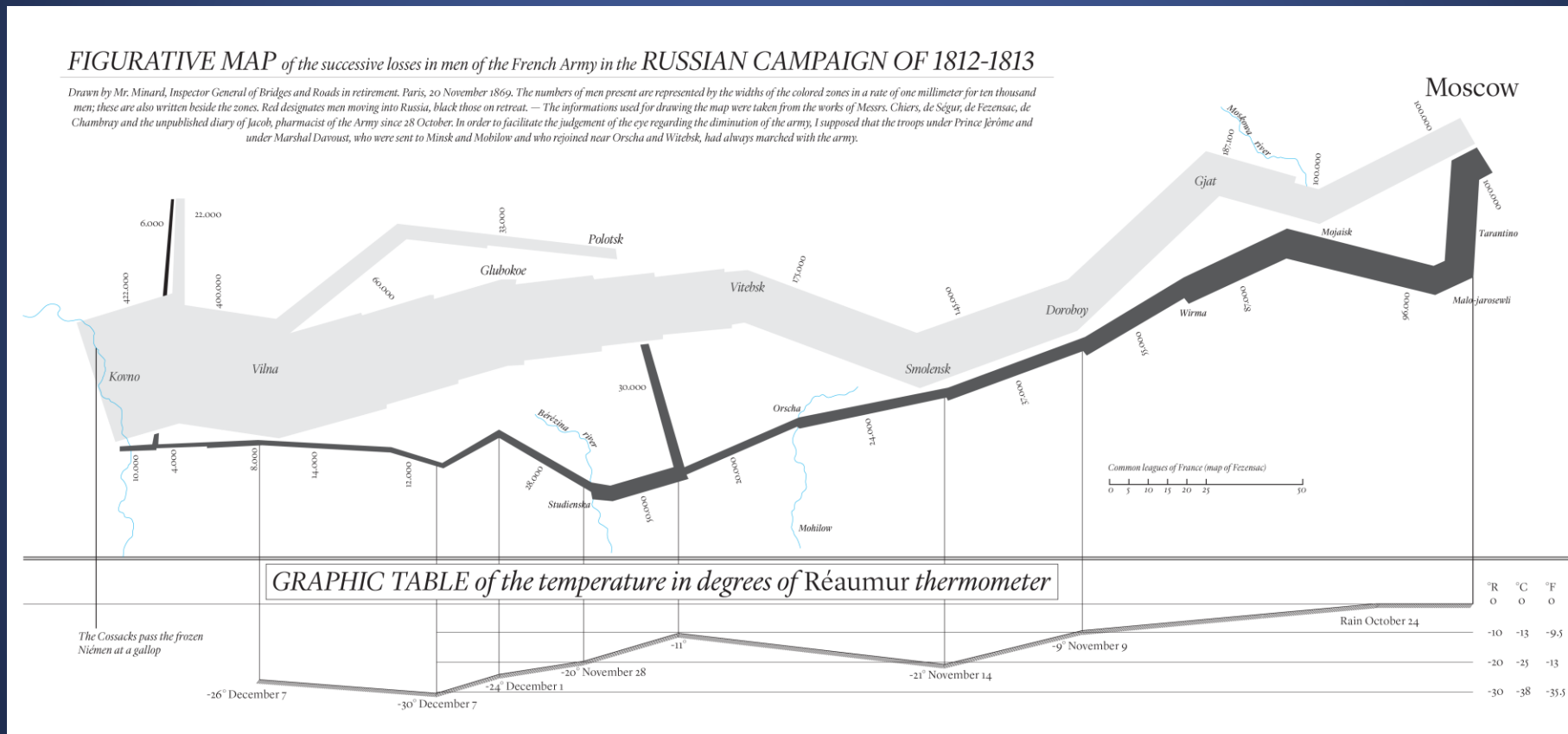


Image source: https://en.wikipedia.org/wiki/Charles_Joseph_Minard

Line size changes as troop strength shrinks as L'Armee de Nord moves into Russia and soldiers drop from battle, disease and cold.  The light line shows movement into Russia, the dark line shows the retreat.

Minard added a line graph at the bottom that shows the relationship between the death rate and the temperature drop during the retreat.
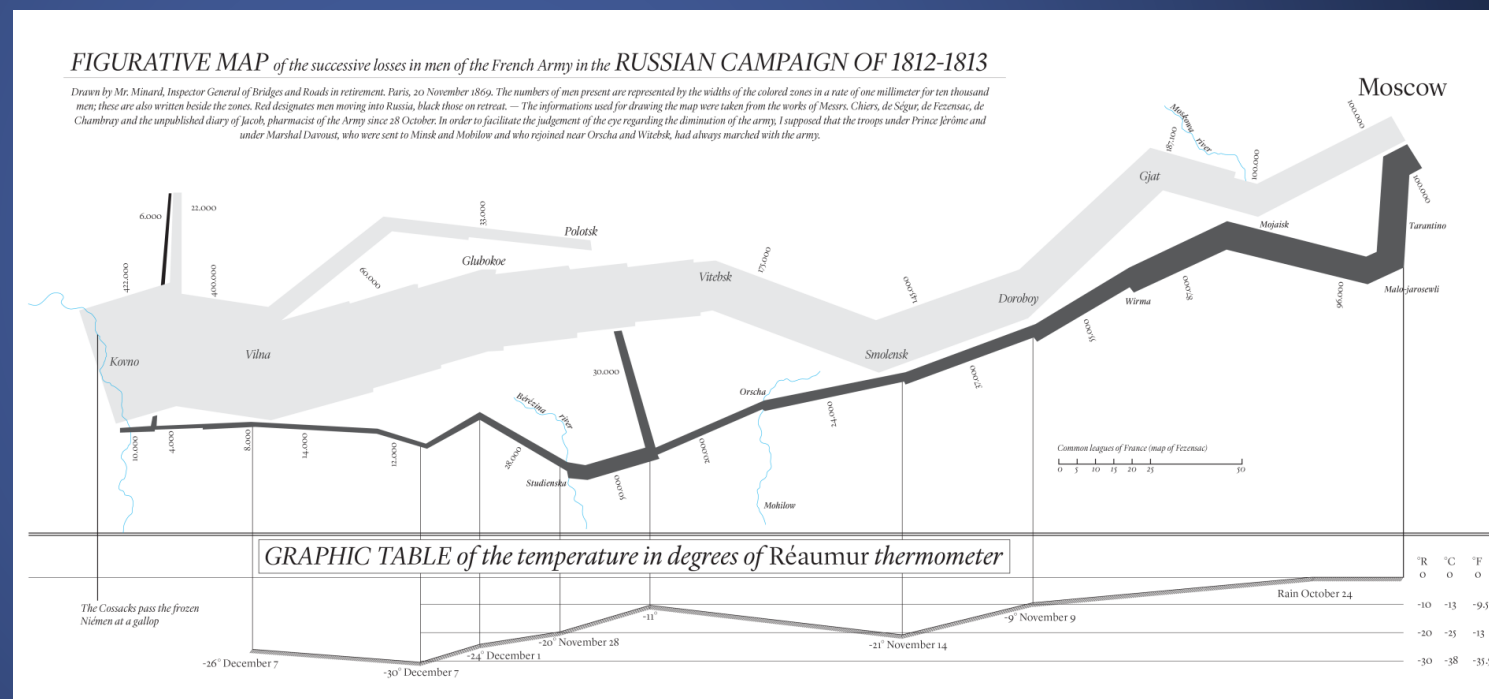


Image source: https://en.wikipedia.org/wiki/Charles_Joseph_Minard

# Sankey diagram software

- Google https://developers.google.com/chart/interactive/docs/gallery/sankey
- R packages: 'riverplot' 'googleVis'
- Tableau https://community.tableau.com/thread/152115
- D3  https://bost.ocks.org/mike/sankey/

# Sankey diagram Tableau

Creating Sankey diagrams in Tableau can be complex and I found the interactivity options were limited.

You can find more here:

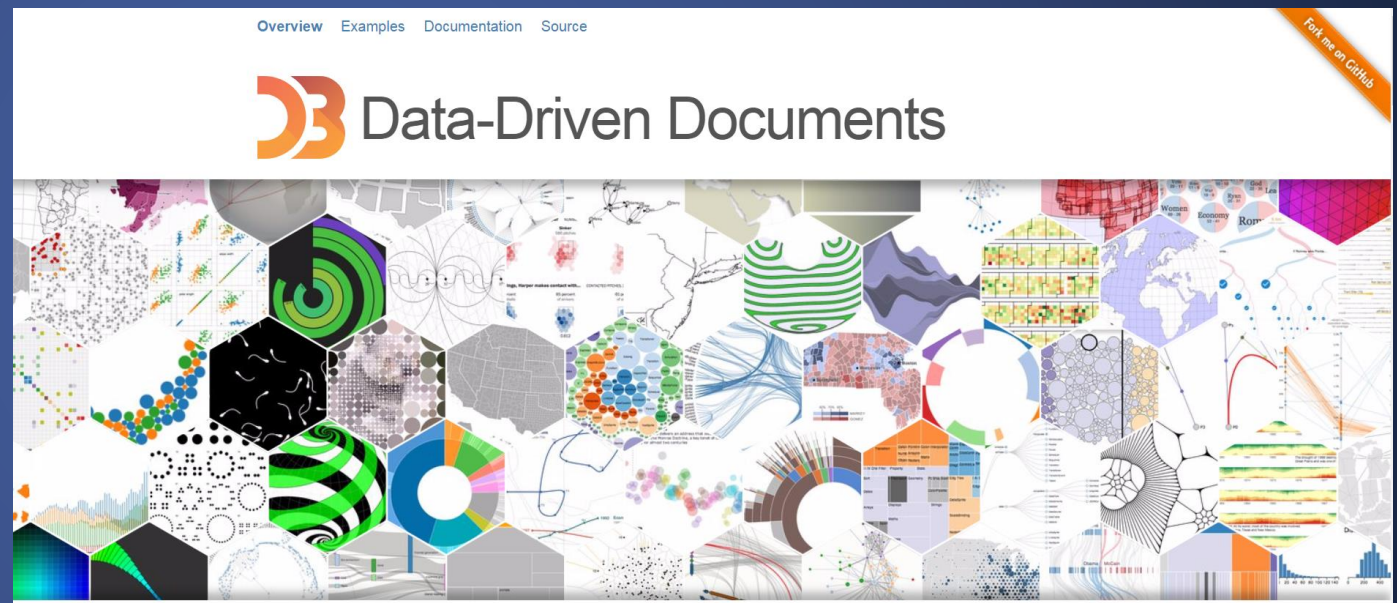https://community.tableau.com/thread/152115
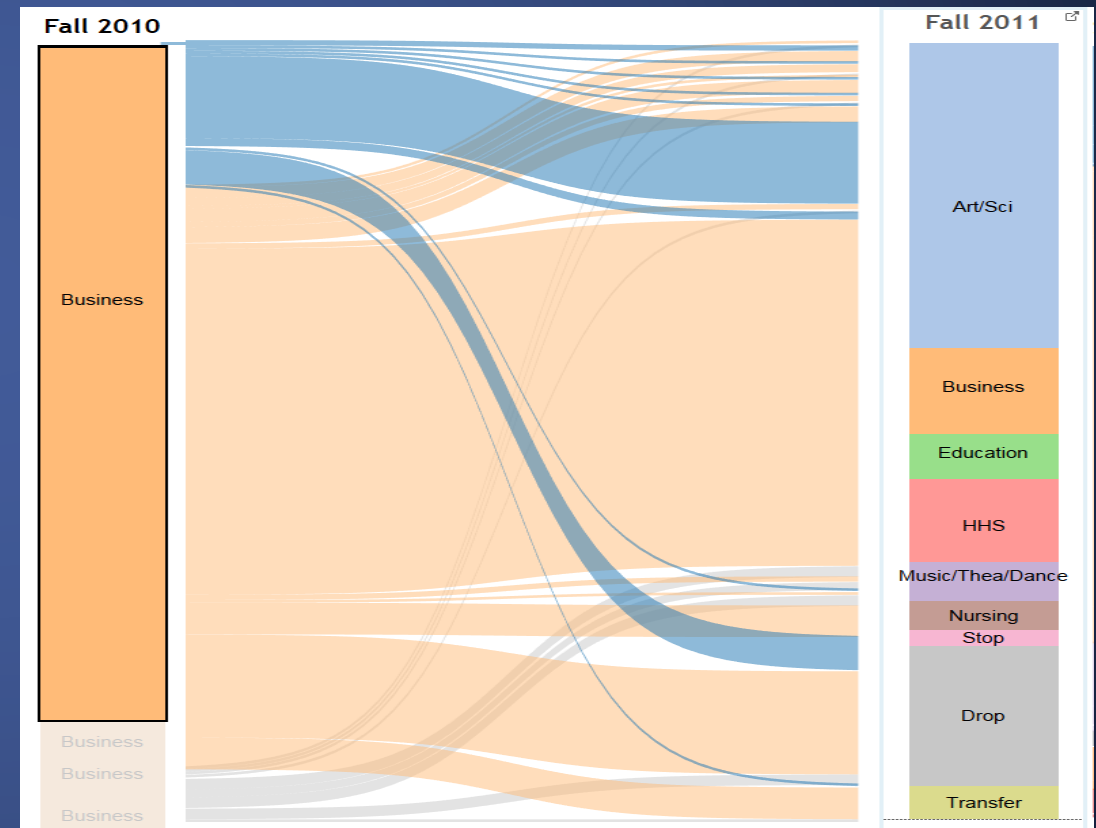
# Sankey diagram D3

D3 is a javascript library designed to ease the process of making interactive web-visualizations

Steep learning curve but well worth the trip

# The Interactive Major Movement Map (Imap)

Five-year IMap for the Bryan School of Business:

• http://ire.uncg.edu/research/Air-2017-Mark-Davenport/IMap_BE_Demo_5.html

Applied—Accepted—Enrolled IMap:

• http://ire.uncg.edu/research/Air-2017-Mark-Davenport/web_index_app.html

# Data setup: Raw

# Step 1: Create Parent/Child pairs

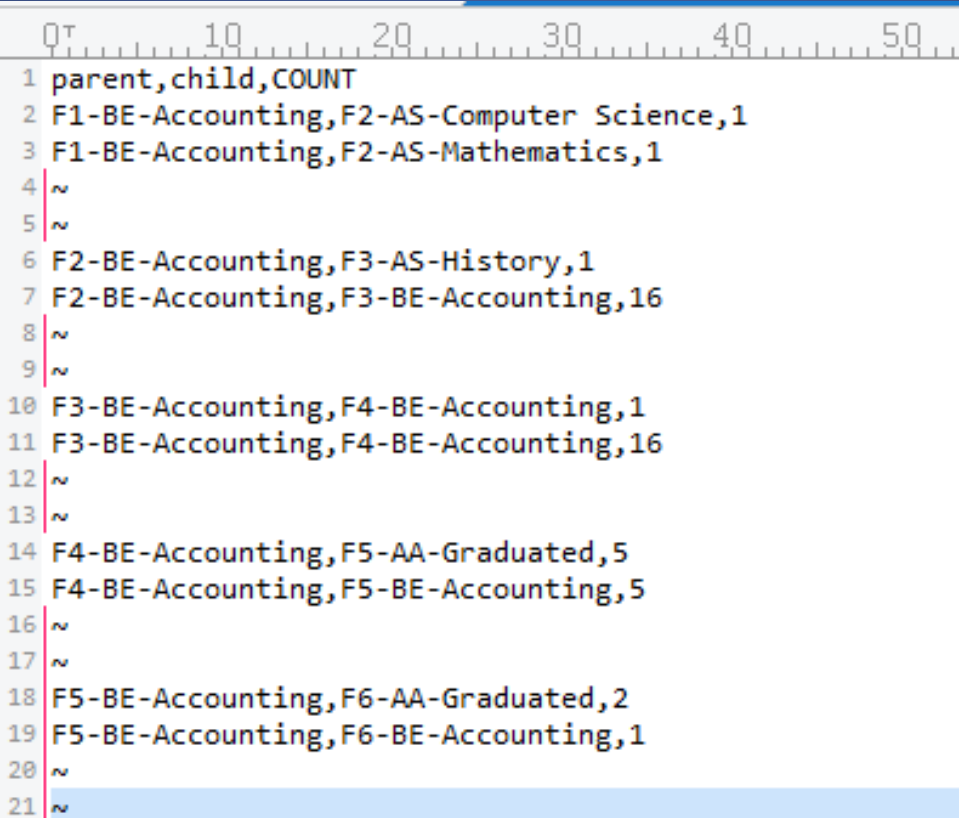| MAJOR_STRING | MAJOR_STR_12 | MAJOR_STR_23 | MAJOR_STR_34 | MAJOR_STR_45 | MAJOR_STR_56 | MAJORN_F1 | MAJORN_F2 |
|---|---|---|---|---|---|---|---|
| PSYCSOCISOCIDDDBADMBADM | PSYCSOCI | SOCISOCI | SOCIDDDD | DDDDBADM | BADMBADM | Psychology | Sociology |
| PSPEPELMELEDELEDGGGGGGGGG | PSPEPELM | PELMELED | ELEDELED | ELEDGGGG | GGGGGGGG | Special Ed - Elem Ed-Pre M... | Elementary Ed-Pre Maj |
| PHLTPSWKDDDDDDDDDDDDDDDD | PHLTPSWK | PSWKDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Pre-Health Studies-Sciences | Pre-Social Work |
| CHEMUNDCPSWKSOWKSOWKGGGG | CHEMUNDC | UNDCPSWK | PSWKSOWK | SOWKSOWK | SOWKGGGG | Chemistry | Undecided on Major |
| DRAMDRAMDRAMDRAMDRAMGGGG | DRAMDRAM | DRAMDRAM | DRAMDRAM | DRAMDRAM | DRAMGGGG | Drama | Drama |
| PELMPELMELEDELEDGGGGGGGG | PELMPELM | PELMELED | ELEDELED | ELEDGGGG | GGGGGGGG | Elementary Ed-Pre Major | Elementary Ed-Pre Maj |
| PELMDDDDDDDDDDDDDDDDDDDD | PELMDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Elementary Ed-Pre Major | Dropped-Stopped-Tra |
| SOCISOCISOCISOCIGGGGGGGG | SOCISOCI | SOCISOCI | SOCISOCI | SOCIGGGG | GGGGGGGG | Sociology | Sociology |
| PNURPHLTDDDDDDDDDDDDDDDD | PNURPHLT | PHLTDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Pre-Nursing | Pre-Health Studies-Sci |
| PSCIPSCIPSCIPSCIGGGGGGGG | PSCIPSCI | PSCIPSCI | PSCIPSCI | PSCIGGGG | GGGGGGGG | Political Science | Political Science |
| BADMDDDDDDDDDDDDDDDDDDDD | BADMDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Business Administration | Dropped-Stopped-Tra |
| UNDCPSYCPSYCPSYCGGGGGGGG | UNDCPSYC | PSYCPSYC | PSYCPSYC | PSYCGGGG | GGGGGGGG | Undecided on Major | Psychology |
| NUTRNUTRNUTRNUTRNUTRGGGG | NUTRNUTR | NUTRNUTR | NUTRNUTR | NUTRNUTR | NUTRGGGG | Nutrition | Nutrition |
| PKINPKINPKINKINEGGGGGGGG | PKINPKIN | PKINPKIN | PKINKINE | KINEGGGG | GGGGGGGG | Pre-Kinesiology | Pre-Kinesiology |
| ACCTACCTHISTHISTHISTGGGG | ACCTACCT | ACCTHIST | HISTHIST | HISTHIST | HISTGGGG | Accounting | Accounting |
| PRDFPRDFPRDFPRDFGGGGGGGG | PRDFPRDF | PRDFPRDF | PRDFPRDF | PRDFGGGG | GGGGGGGG | Professions in Deafness | Professions in Deafnes |
| CMPSCMPSCMPSCMPSBIOLGGGG | CMPSCMPS | CMPSCMPS | CMPSCMPS | CMPSBIOL | BIOLGGGG | Computer Science | Computer Science |
| PHLTBIOLBIOLDDDDDDDDDDDD | PHLTBIOL | BIOLBIOL | BIOLDDDD | DDDDDDDD | DDDDDDDD | Pre-Health Studies-Sciences | Biology |
| PELMBIOLBIOLDDDDDDDDDDDD | PELMBIOL | BIOLBIOL | BIOLDDDD | DDDDDDDD | DDDDDDDD | Elementary Ed-Pre Major | Biology |
| PHLTPHLTHDFSHDFSGGGGGGGG | PHLTPHLT | PHLTHDFS | HDFSHDFS | HDFSGGGG | GGGGGGGG | Pre-Health Studies-Sciences | Pre-Health Studies-Sci |
| DRAMUNDCANTHANTHGGGGGGGG | DRAMUNDC | UNDCANTH | ANTHANTH | ANTHGGGG | GGGGGGGG | Drama | Undecided on Major |
| CHEMBCHEKINEKINEKINEGGGG | CHEMBCHE | BCHEKINE | KINEKINE | KINEKINE | KINEGGGG | Chemistry | Biochemistry |
| PSYCDDDDDDDDDDDDDDDDDDDD | PSYCDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Psychology | Dropped-Stopped-Tra |
| CHEMSOCIDDDDDDDDDDDDDDDD | CHEMSOCI | SOCIDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Chemistry | Sociology |
| BADMECONECONECONGGGGGGGG | BADMECON | ECONECON | ECONECON | ECONGGGG | GGGGGGGG | Business Administration | Economics |

# Step 2: Sort Parent/Child pairs

| MAJOR_STRING | MAJOR_STR_12 | MAJOR_STR_23 | MAJOR_STR_34 | MAJOR_STR_45 | MAJOR_STR_56 | MAJORN_F1 | MAJORN_F2 | |
|---|---|---|---|---|---|---|---|---|
| AADSAADSAADSAADSDDDDDDDD | AADSAADS | AADSAADS | AADSAADS | AADSDDDD | DDDDDDDD | Afr Amn-Afr Diaspora Studies | Afr Amn-Afr Diaspora Studies | A |
| ACCTACCTACCTACCTACCTDDDD | ACCTACCT | ACCTACCT | ACCTACCT | ACCTACCT | ACCTDDDD | Accounting | Accounting | A |
| ACCTACCTACCTACCTACCTGGGG | ACCTACCT | ACCTACCT | ACCTACCT | ACCTACCT | ACCTGGGG | Accounting | Accounting | A |
| ACCTACCTACCTACCTGGGGGGGG | ACCTACCT | ACCTACCT | ACCTACCT | ACCTGGGG | GGGGGGGG | Accounting | Accounting | A |
| ACCTACCTACCTDDDDDDDDDDDD | ACCTACCT | ACCTACCT | ACCTDDDD | DDDDDDDD | DDDDDDDD | Accounting | Accounting | A |
| ACCTACCTACCTENTRENTRGGGG | ACCTACCT | ACCTACCT | ACCTENTR | ENTRENTR | ENTRGGGG | Accounting | Accounting | A |
| ACCTACCTACCTFINCBADMGGGG | ACCTACCT | ACCTACCT | ACCTFINC | FINCBADM | BADMGGGG | Accounting | Accounting | A |
| ACCTACCTACCTFINCGGGGGGGG | ACCTACCT | ACCTACCT | ACCTFINC | FINCGGGG | GGGGGGGG | Accounting | Accounting | A |
| ACCTACCTARTARTARTGGGG | ACCTACCT | ACCTART | ARTART | ARTART | ARTGGGG | Accounting | Accounting | A |
| ACCTACCTBADMBADMBADMGGGG | ACCTACCT | ACCTBADM | BADMBADM | BADMBADM | BADMGGGG | Accounting | Accounting | B |
| ACCTACCTBADMBADMGGGGGGGG | ACCTACCT | ACCTBADM | BADMBADM | BADMGGGG | GGGGGGGG | Accounting | Accounting | B |
| ACCTACCTDDDDDDDDDDDDDDDD | ACCTACCT | ACCTDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Accounting | Accounting | D |
| ACCTACCTECONECONECONGGGG | ACCTACCT | ACCTECON | ECONECON | ECONECON | ECONGGGG | Accounting | Accounting | E |
| ACCTACCTFINCFINCFINCDDDD | ACCTACCT | ACCTFINC | FINCFINC | FINCFINC | FINCDDDD | Accounting | Accounting | F |
| ACCTACCTHISTHISTHISTGGGG | ACCTACCT | ACCTHIST | HISTHIST | HISTHIST | HISTGGGG | Accounting | Accounting | H |
| ACCTACCTNUTRDDDDFINCFINC | ACCTACCT | ACCTNUTR | NUTRDDDD | DDDDFINC | FINCFINC | Accounting | Accounting | N |
| ACCTBADMBADMBADMGGGGGGGG | ACCTBADM | BADMBADM | BADMBADM | BADMGGGG | GGGGGGGG | Accounting | Business Administration | B |
| ACCTBADMSOCISOCIGGGGGGGG | ACCTBADM | BADMSOCI | SOCISOCI | SOCIGGGG | GGGGGGGG | Accounting | Business Administration | S |
| ACCTCMPSCMPSCMPSISSCISSC | ACCTCMPS | CMPSCMPS | CMPSCMPS | CMPSISSC | ISSCISSC | Accounting | Computer Science | C |
| ACCTDDDDDDDDDDDDACCTACCT | ACCTDDDD | DDDDDDDD | DDDDDDDD | DDDDACCT | ACCTACCT | Accounting | Dropped-Stopped-Transferr... | D |
| ACCTDDDDDDDDDDDDDDDDDDDD | ACCTDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Accounting | Dropped-Stopped-Transferr... | D |
| ACCTFINCFINCINTBGGGGGGGG | ACCTFINC | FINCFINC | FINCINTB | INTBGGGG | GGGGGGGG | Accounting | Finance | F |
| ACCTKINEACCTACCTACCTDDDD | ACCTKINE | KINEACCT | ACCTACCT | ACCTACCT | ACCTDDDD | Accounting | Kinesiology | A |
| ACCTMATHMATHDDDDDDDDGGGG | ACCTMATH | MATHMATH | MATHDDDD | DDDDDDDD | DDDDGGGG | Accounting | Mathematics | M |
| ACCTUNDCDDDDDDDDDDDDDDDD | ACCTUNDC | UNDCDDDD | DDDDDDDD | DDDDDDDD | DDDDDDDD | Accounting | Undecided on Major | D |

# Step 3: Make labels, aggregate, and count

| PARENT | CHILD | MAJOR_STR_12 | F1_F2_COUNT |
|---|---|---|---|
| BE--Accounting | BE--Accounting | ACCTACCT | 24 |
| BE--Accounting | BE--Business Administration | ACCTBADM | 2 |
| BE--Accounting | AS--Computer Science | ACCTCMPS | 1 |
| BE--Accounting | ZZ--Dropped-Stopped-Transferred | ACCTDDDD | 12 |
| BE--Accounting | BE--Finance | ACCTFINC | 1 |
| BE--Accounting | HH--Kinesiology | ACCTKINE | 1 |
| BE--Accounting | AS--Mathematics | ACCTMATH | 1 |
| BE--Accounting | AS--Undecided on Major | ACCTUNDC | 1 |
| BE--Business Administration | BE--Accounting | BADMACCT | 5 |
| BE--Business Administration | BE--Business Administration | BADMBADM | 48 |
| BE--Business Administration | AS--Biology | BADMBIOL | 1 |
| BE--Business Administration | AS--Computer Science | BADMCMPS | 1 |
| BE--Business Administration | VP--Dance | BADMDANC | 1 |
| BE--Business Administration | ZZ--Dropped-Stopped-Transferred | BADMDDDD | 31 |
| BE--Business Administration | BE--Economics | BADMECON | 2 |

# Steps 1-3 are repeated for every parent/child pair.  For six terms, it is run five times

```
    QT        10        20        30        40        50
 1  parent,child,COUNT
 2  F1-BE-Accounting,F2-AS-Computer Science,1
 3  F1-BE-Accounting,F2-AS-Mathematics,1
 4  ~
 5  ~
 6  F2-BE-Accounting,F3-AS-History,1
 7  F2-BE-Accounting,F3-BE-Accounting,16
 8  ~
 9  ~
10  F3-BE-Accounting,F4-BE-Accounting,1
11  F3-BE-Accounting,F4-BE-Accounting,16
12  ~
13  ~
14  F4-BE-Accounting,F5-AA-Graduated,5
15  F4-BE-Accounting,F5-BE-Accounting,5
16  ~
17  ~
18  F5-BE-Accounting,F6-AA-Graduated,2
19  F5-BE-Accounting,F6-BE-Accounting,1
20  ~
21  ~
```
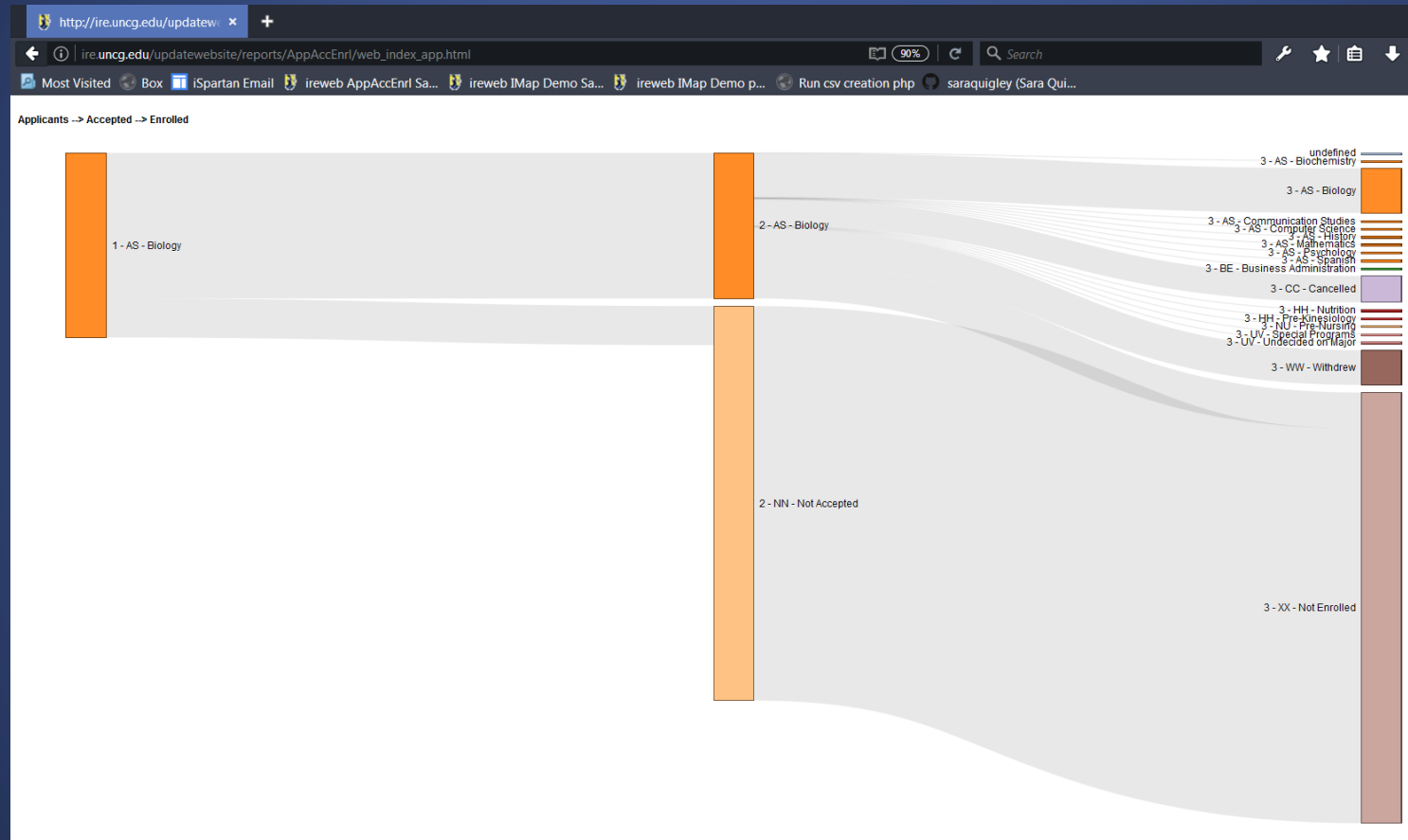
# Summary of data steps
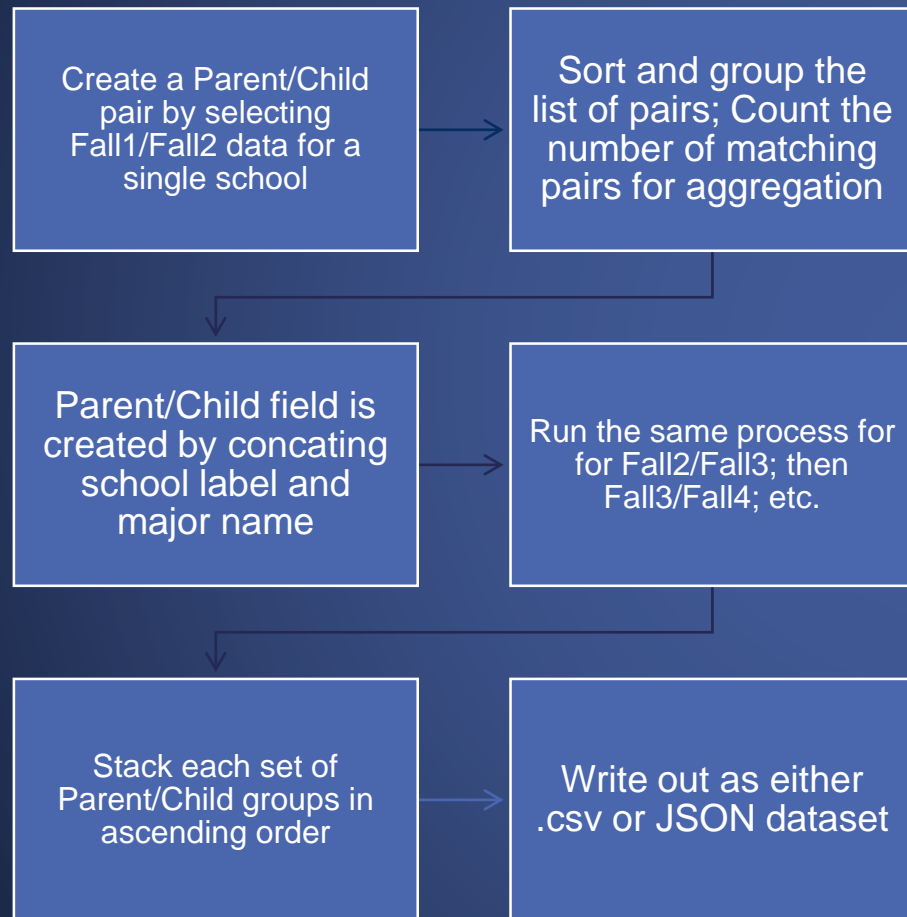


6/27/2017

# The IMap
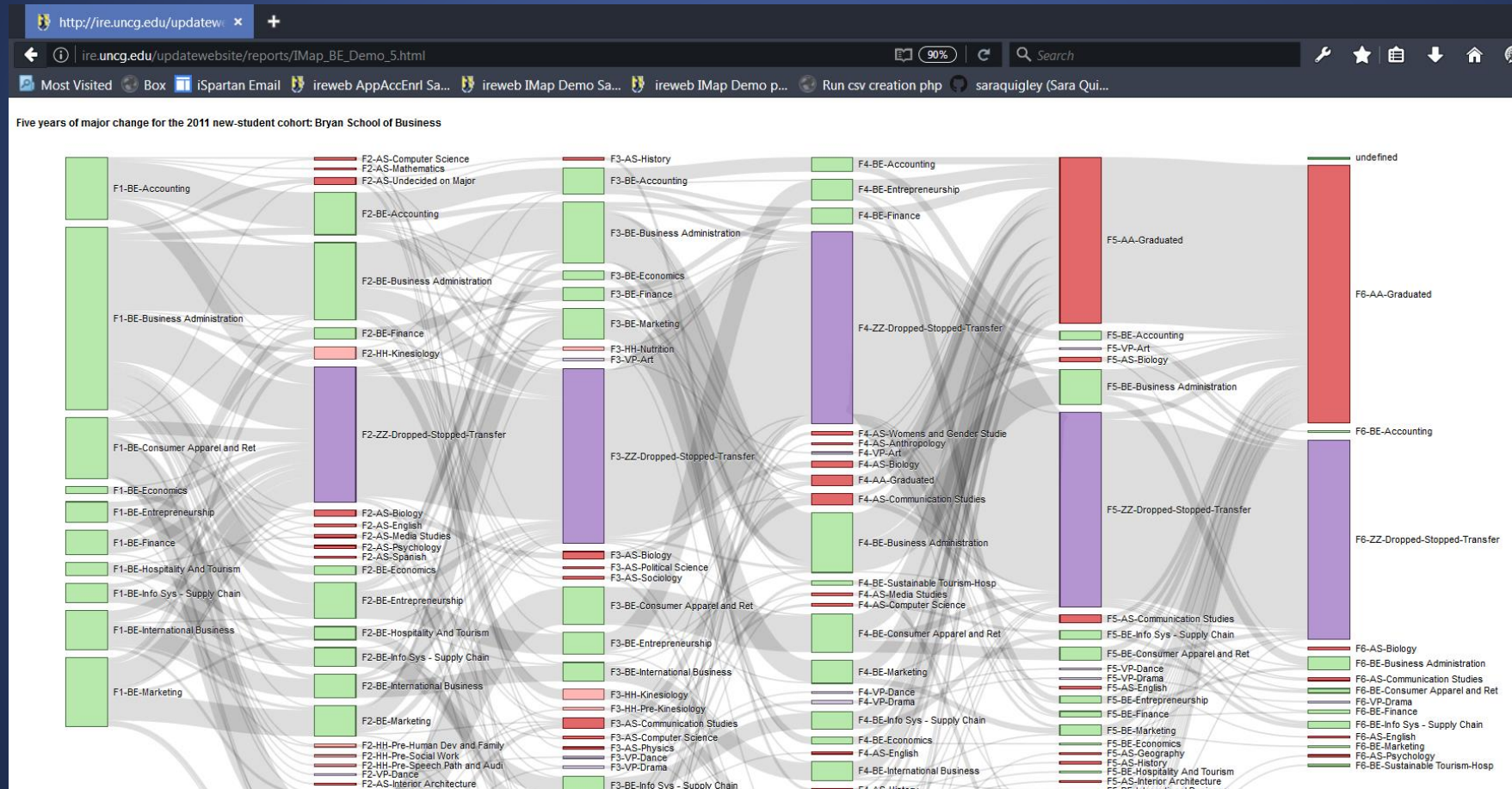
# Another example: Applied-Accepted-Enrolled IMap

# Future improvements

- Fully functional webpage that allows schools and departments to drill down to their own students

- PHP/MySql functionality that runs the entire process from raw data at the click of a button

- Programs to be posted on my GitHub site when complete

# Acknowledgements

Much of the code used in this project was produced by others and posted on GitHub.  My gratitude to:


Sara Quigley -- https://github.com/saraquigley


Mike Bostock – https://d3js.org/  Creator of D3