

Data Visualization & Exploration

Lecture 2

Angela Zoss

Data Visualization Coordinator
Data and Visualization Services

Eric Monson

Data Visualization Analyst
Data and Visualization Services

1 simple dataset: Which is the best chart? And why?

Even with just 20 values, this is a difficult question to answer.

Andy Cotgreave, Tableau
@acotgreave



0:00 / 5:29



<https://www.youtube.com/watch?v=AuJFuEq-qD8>

COMPARISON

Supporting successful comparisons:

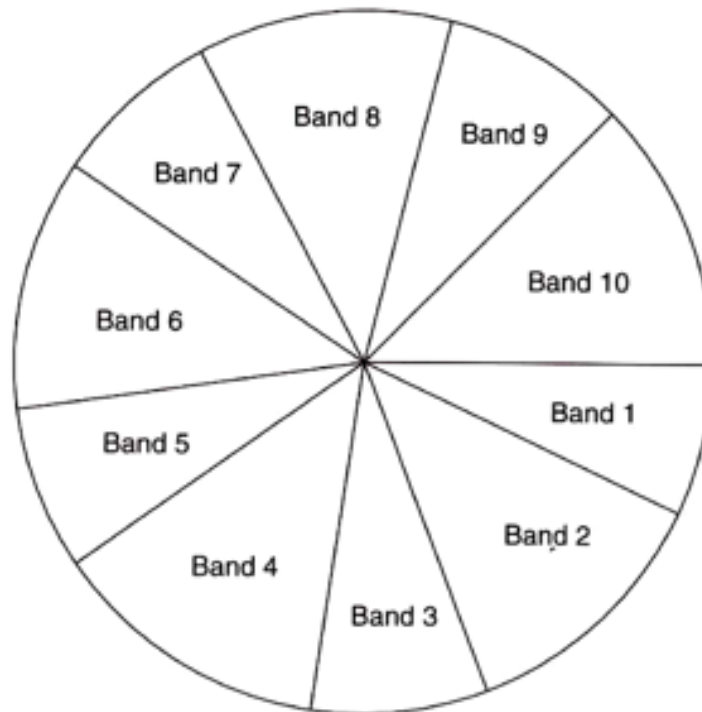
1. Data properties
2. Visualization properties
3. Expected tasks
4. User skills
5. Design context

See also: <http://complexdiagrams.com/4pillars>

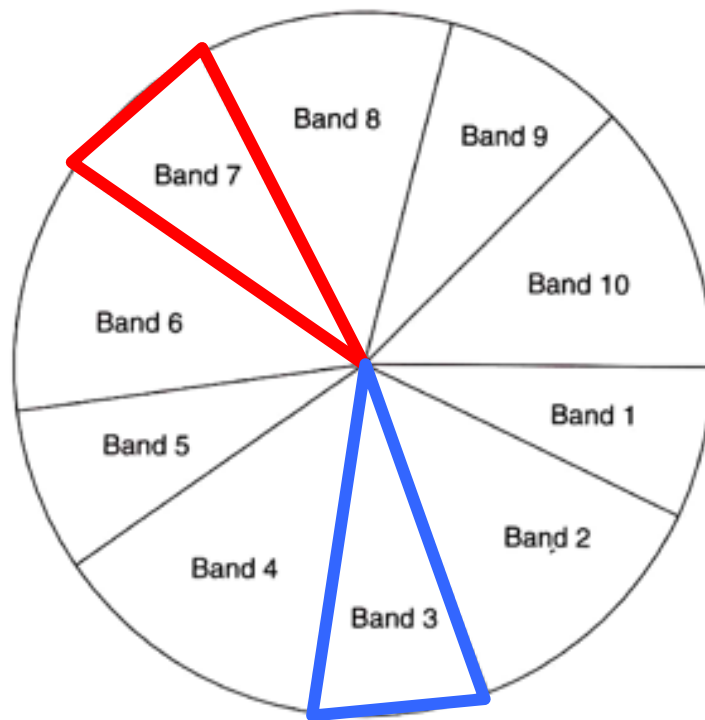
Supporting successful comparisons:

1. Data properties
2. Visualization properties
3. Expected tasks
4. User skills
5. Design context

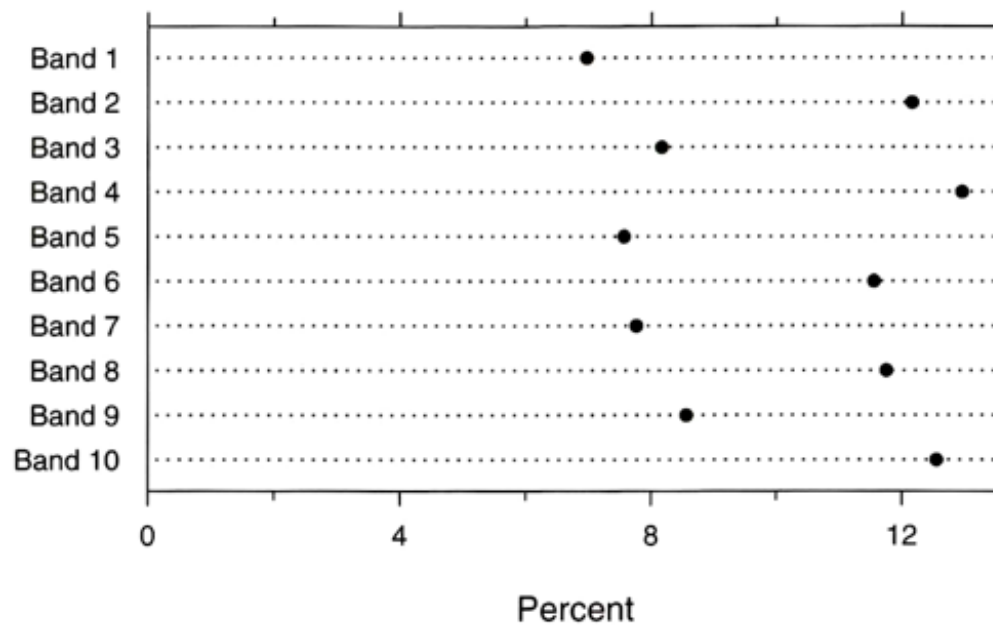
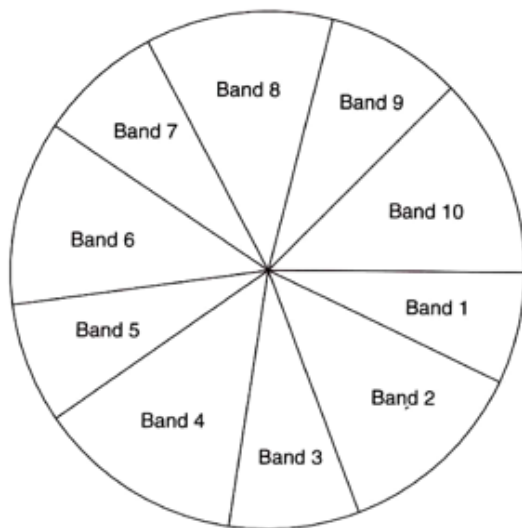
See also: <http://complexdiagrams.com/4pillars>



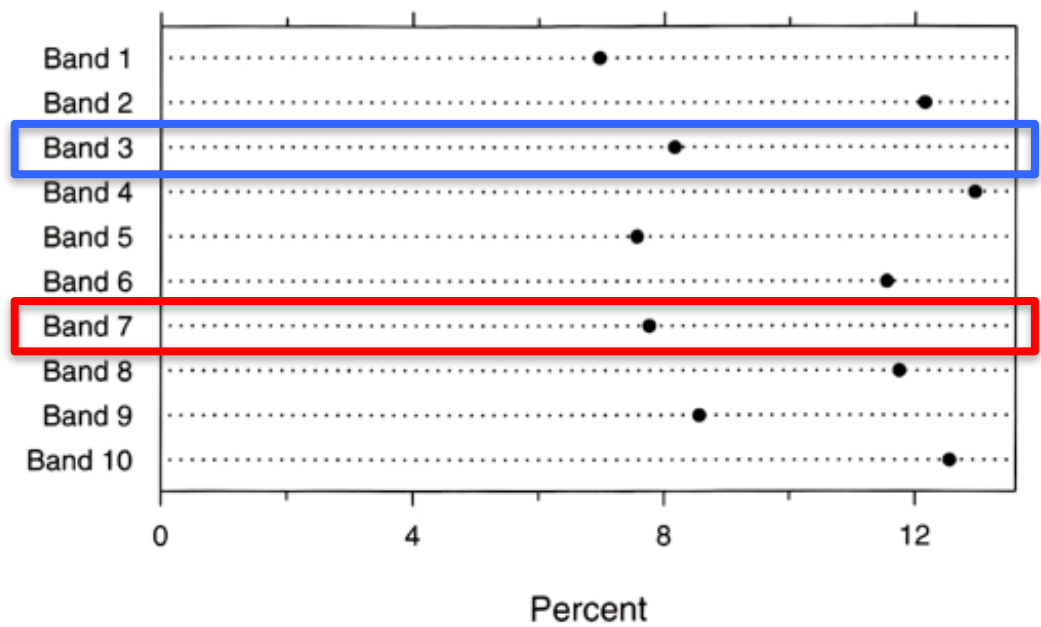
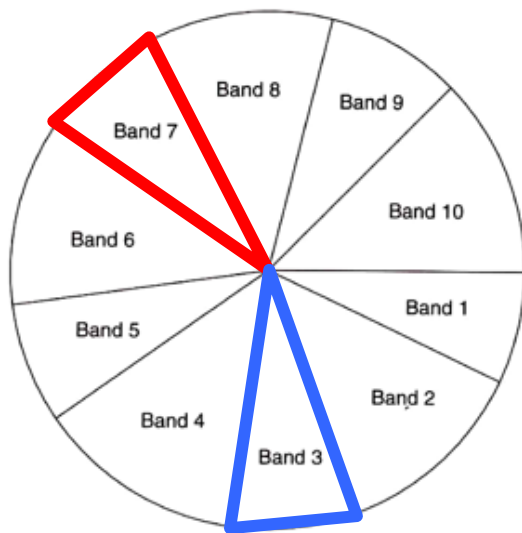
Cleveland (1994)



Cleveland (1994)

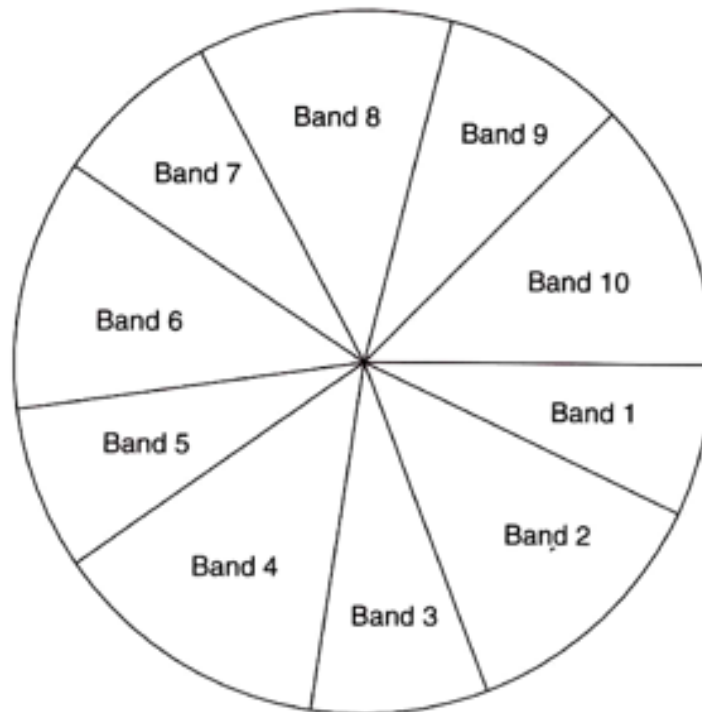


Cleveland (1994)



Cleveland (1994)

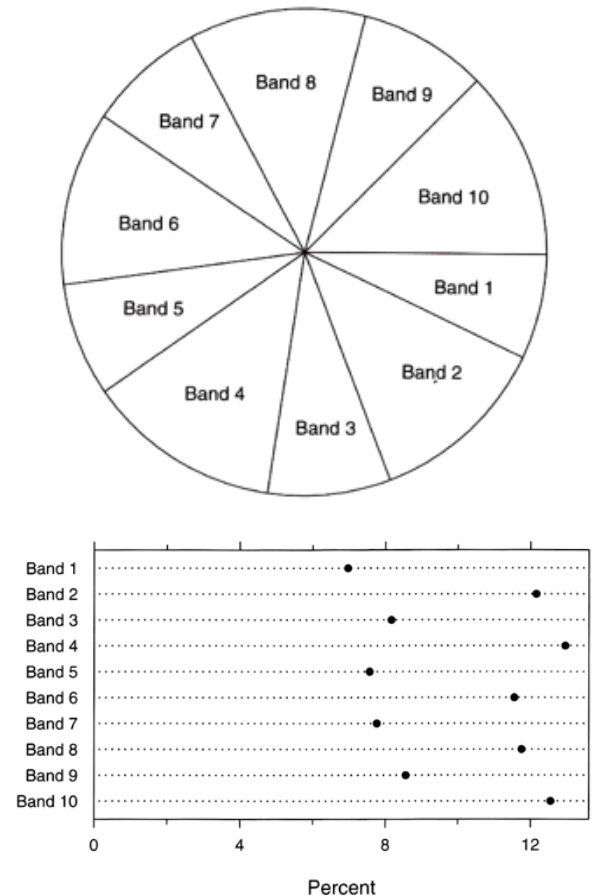
DATA PROPERTIES



Cleveland (1994)

Important features

- 10 numerical measurements
- Ordered series of bands
- Alternating magnitudes
- Small differences within each subgroup
- Values in larger group about 50% larger than values in smaller group



Cleveland (1994)

Important features, general

- Data types
do you have mostly numerical or categorical data?
- Minimum and maximum values
how large a range of values do you have?
- Distribution of values
do your data vary a lot? in a regular way?
- Number of measurements
do you have a lot of data? not much at all?

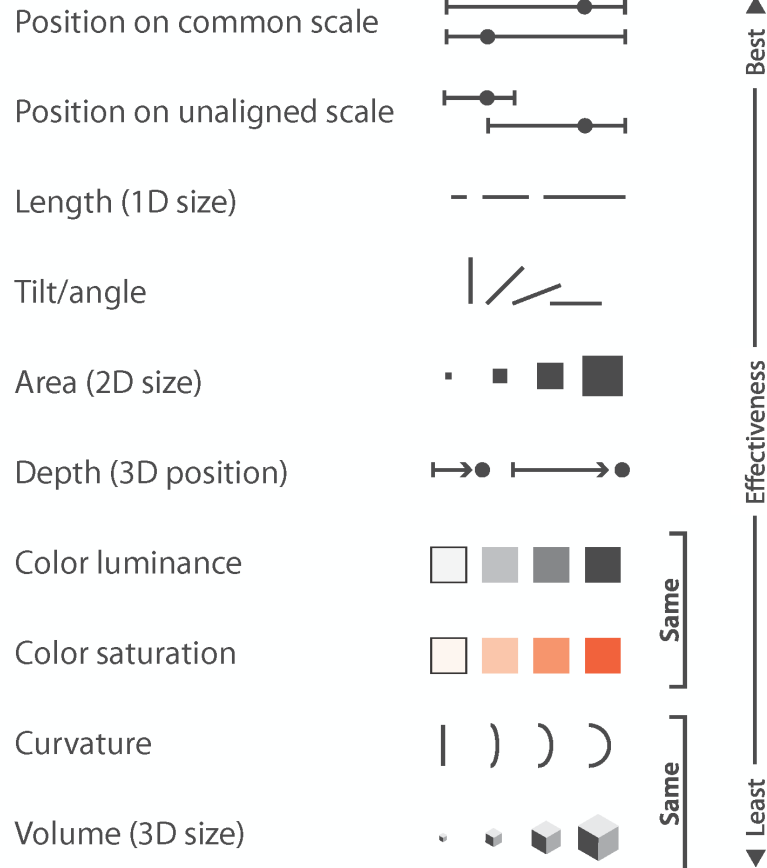
Data properties vs. Visualization properties

- Which data types match with which charts?
- Do certain patterns in the data cause problems with certain charts?

VISUALIZATION PROPERTIES

Different data types, different encodings

Numerical



Categorical



- **expressiveness principle**
 - match channel and data characteristics
- **effectiveness principle**
 - encode most important attributes with highest ranked channels

Tamara Munzner

<http://www.cs.ubc.ca/~tmm/talks/minicourse14/vad16act.pdf>

Encode the most important variables with position

Position is everything.
Color is difficult.

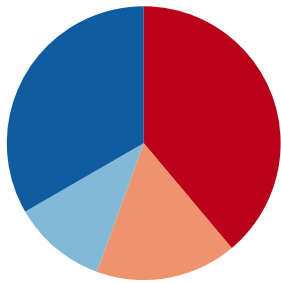
-@moritz_stefaner

Different chart types, different encodings

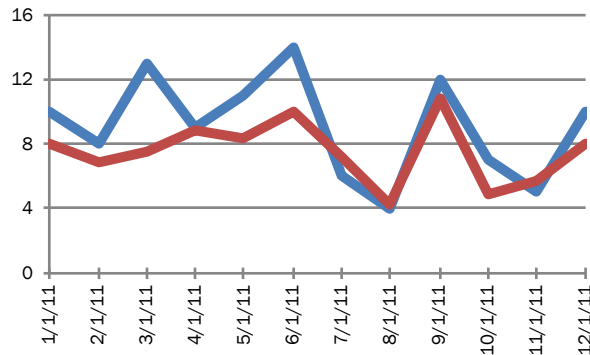
	Categorical variables	Encoded as...	Numerical variables	Encoded as...	Good for...
Pie chart	1	Color	1	Angle	Simple proportions
Line chart	0 or 1	Color	2 (including date)	Position	Trends over time
Bar chart	1 or 2	Position, color	1	Length	Broad audiences, precise comparisons
Scatter plot	0 or 1	Color	2	Position	Correlations
Bubble chart	0 or 1	Color	3	Position, area	Extra variables
Heatmap	0, 1, or 2	Position	1, 2, or 3	Color, position	Comparing categorical variables

Zoss, Angela M. “Designing Public Visualizations of Library Data.” In *Data Visualization: A Guide to Visual Storytelling for Librarians*, edited by Lauren Magnuson. Lanham, MD: Rowman & Littlefield Publishers, Inc., forthcoming. <https://iu.box.com/DesigningPublicVis>

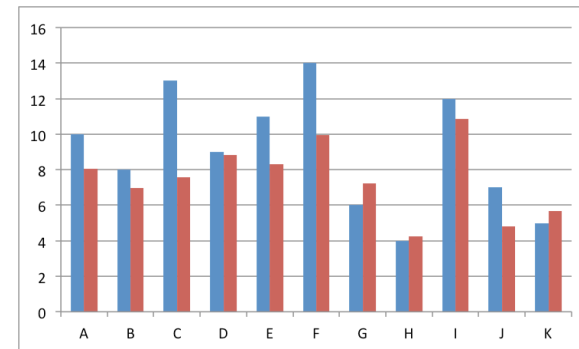
Different chart types, different encodings



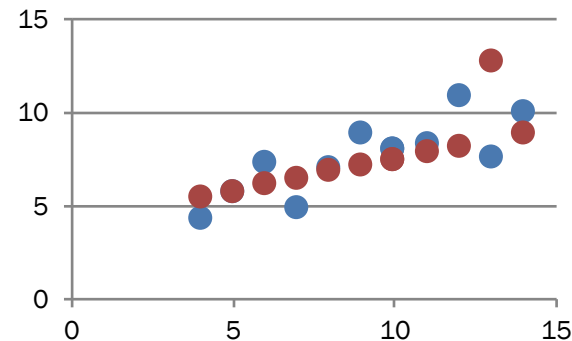
1 categorical: color/position
1 numerical: area



1 date/numerical: x position
1 numerical: y position
1 categorical: color (optional)



1 categorical: x (or y) position
1 numerical: length
1 categorical: color (optional)

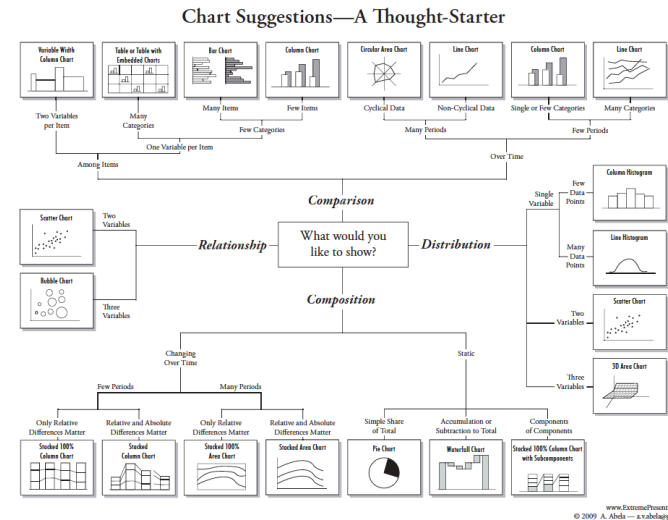


2 numerical: x & y position
1 categorical: color (optional)
1 numerical: size (optional)

EXPECTED TASKS

Possible chart viewing tasks

- Comparison
- Distribution
- Composition
- Relationship



http://extremerepresentation.typepad.com/blog/2006/09/choosing_a_good.html

(N.B.: will likely depend on what is *interesting* about the data.)

Charts for comparison

Based on position on an axis



Bar Chart



Box & Whisker Plot



Bubble Chart



Bullet Graph



Histogram



Line Graph



Marimekko Chart



Multi-set Bar Chart



Nightingale Rose Chart



Population Pyramid



Radar Chart



Radial Bar Chart



Radial Column Chart



Span Chart



Stacked Area Graph



Stacked Bar Graph

No Axis



Chord Diagram



Choropleth Map



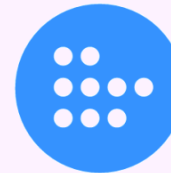
Donut Chart



Dot Matrix Chart



Parallel Sets



Pictogram Chart



Pie Chart



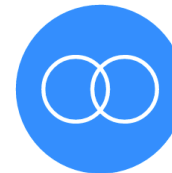
Proportional Area Chart



Tally Chart



Treemap



Venn Diagram

<http://www.datavizcatalogue.com>

Charts for distribution

Visualisation methods that display frequency, how data spread out over an interval or is grouped.



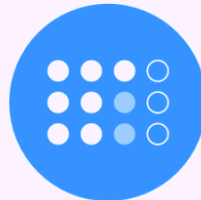
Box & Whisker Plot



Bubble Chart



Density Plot



Dot Matrix Chart



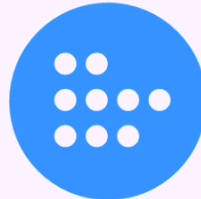
Histogram



Multi-set Bar Chart



Parallel Sets



Pictogram Chart



Stem & Leaf Plot



Tally Chart

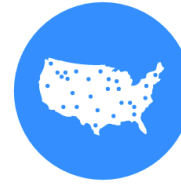


Timeline



Violin Plot

Distribution shown geographically



Dot Map



Connection Map



Flow Map

Distribution shown over age and sex in a population



Population Pyramid

Distribution in a body of text



Word Cloud

<http://www.datavizcatalogue.com>

Charts for composition

Proportions between values



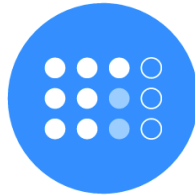
Bubble Chart



Bubble Map



Circle Packing



Dot Matrix Chart



Nightingale Rose Chart



Proportional Area Chart



Stacked Bar Graph



Word Cloud

Proportions in parts-to-a-whole



Donut Chart



Marimekko Chart



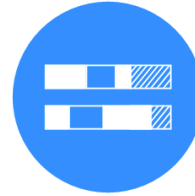
Parallel Sets



Pie Chart



Sankey Diagram



Stacked Bar Graph



Treemap



Sunburst Diagram

Charts for relationship



Arc Diagram



Bar Chart



Brainstorm



Bubble Chart



Radial Bar Chart



Scatterplot



Stacked Area Graph



Stacked Bar Graph



Candlestick Chart



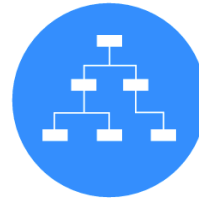
Chord Diagram



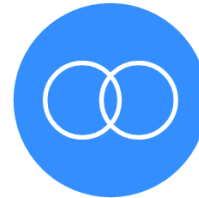
Connection Map



Line Graph



Tree Diagram



Venn Diagram



Marimekko Chart



Network Diagram



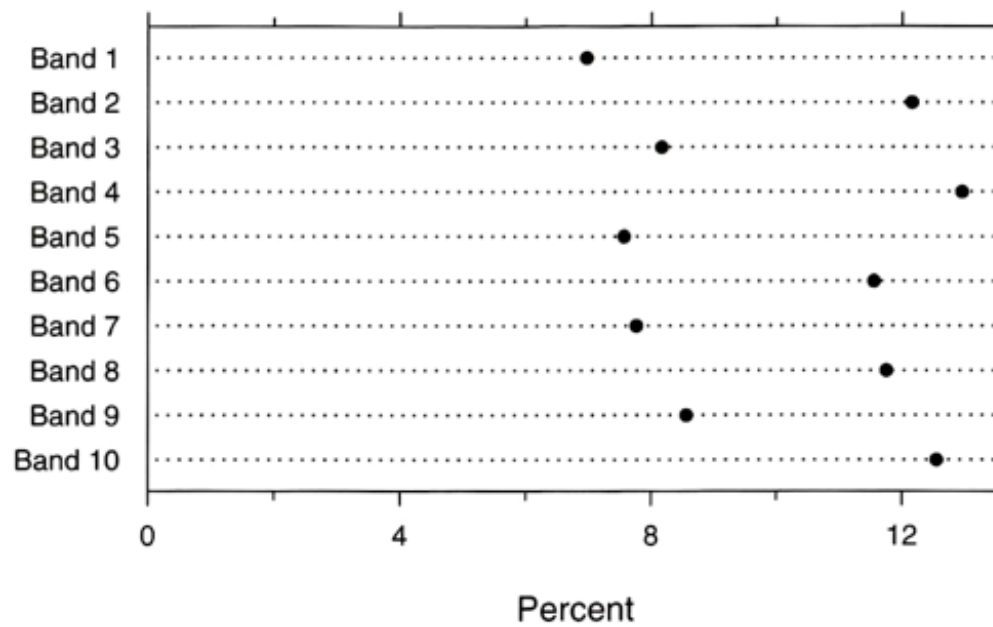
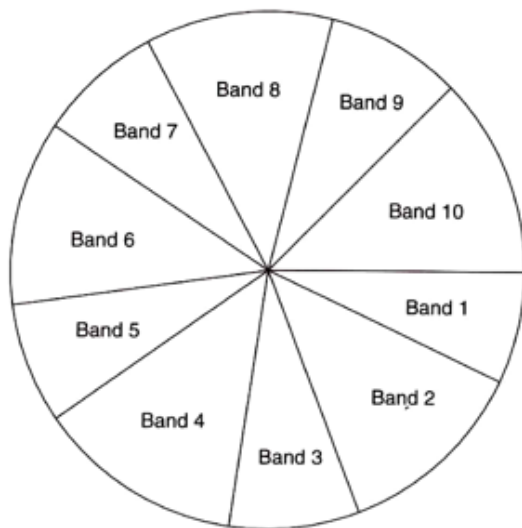
Non-ribbon Chord Diagram



Radar Chart

Choose a chart that:

- Fits the **data types** you have
- Supports the **task** you want to help with
- Uses **position** for the variable you care most about



Cleveland (1994)

MORE ABOUT VISUALIZATION

Data Visualization LibGuides

- Data visualization:
<http://guides.library.duke.edu/datavis>
- Visualization types:
http://guides.library.duke.edu/vis_types
- Top 10 dos and don'ts for charts and graphs:
<http://guides.library.duke.edu/topten>
- Visual communication:
<http://guides.library.duke.edu/visualcomm>

Visualization Friday Forum

- Every Friday during the semester
- Noon in LSRC D106
- Free lunch (sssh!)



WHERE CAN I GO FOR HELP?

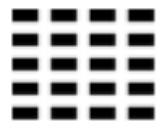
Data and Visualization Services

- Computing cluster (The Edge)
- Walk-in consultations
- Data collections
- Workshops
- Online instructional materials



<http://library.duke.edu/data>

Support Areas



Data Sources



Data Management



Data Cleaning



Data Analysis



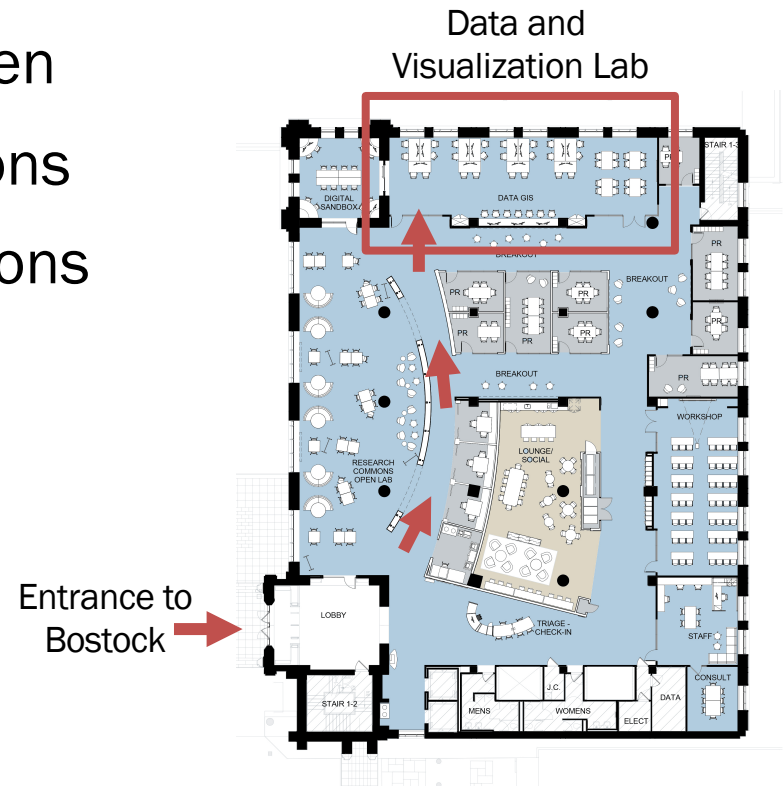
Mapping and GIS



Data Visualization

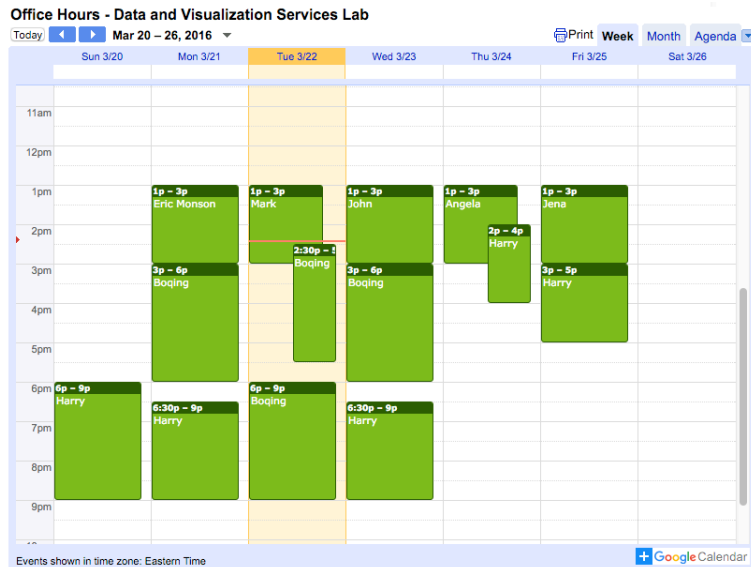
Brandaleone Family Lab for Data and Visualization Services

- The Edge (1st floor of Bostock)
- Open whenever the library is open
- 12 high-powered Dell workstations
- 3 Bloomberg financial workstations
- Software for data analysis, GIS, and visualization



<http://library.duke.edu/data/about/lab>

Walk-in Consulting



...or email
askdata@duke.edu
for an appointment

<http://library.duke.edu/data/about/schedule>

Workshops

- Typically toward the beginning of the semester
- Covering: visualization, data processing, GIS/mapping
- 1-2 hours, often hands-on

<http://library.duke.edu/data/news>



For announcements, sign up for our listserv:

<https://lists.duke.edu/sympa/subscribe/dvs-announce>

askdata@duke.edu

[@duke_vis](#), [@duke_data](#)

QUESTIONS?