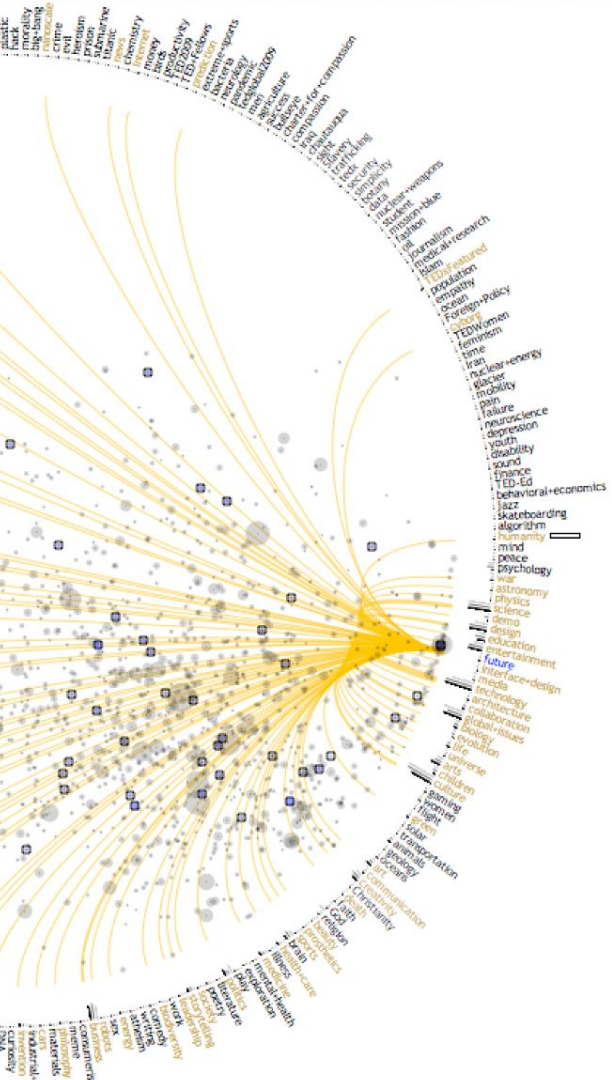
We will get started at 3:10

Relax

Midterms have you stressed out? Here's 10 minutes to relax. I can answer questions now but I'll be free after class as well.



Patterns

A Samuel Pottinger
Stat 198: IDSV
Mar 5, 2025

Today

> Variables: dimensions and measures.

Patterns: a sampling of options in increasing density.

Group activity: choosing a pattern for a problem.

Tasks and domains: a look at contextualizing patterns by understanding users.

What do these terms mean?

- Observation
- Value
- Variable
- Dimension
- Measure

Observations

Variables
The column names.

Values
A value seen inside a cell.

Observations
Rows in a table.

Age	Income	Region
20	100,000	West
31	140,000	East
42	160,000	West

Journal of Statistical Software
August 2014, Volume 59, Issue 10. *http://www.jstatsoft.org/*

Tidy Data

Hadley Wickham
RStudio

Variables

Dimensions
Variables by which observations are divided or segmented.

Measures
Numeric values which are encoded in visual channels.

tableau
from Salesforce

Today

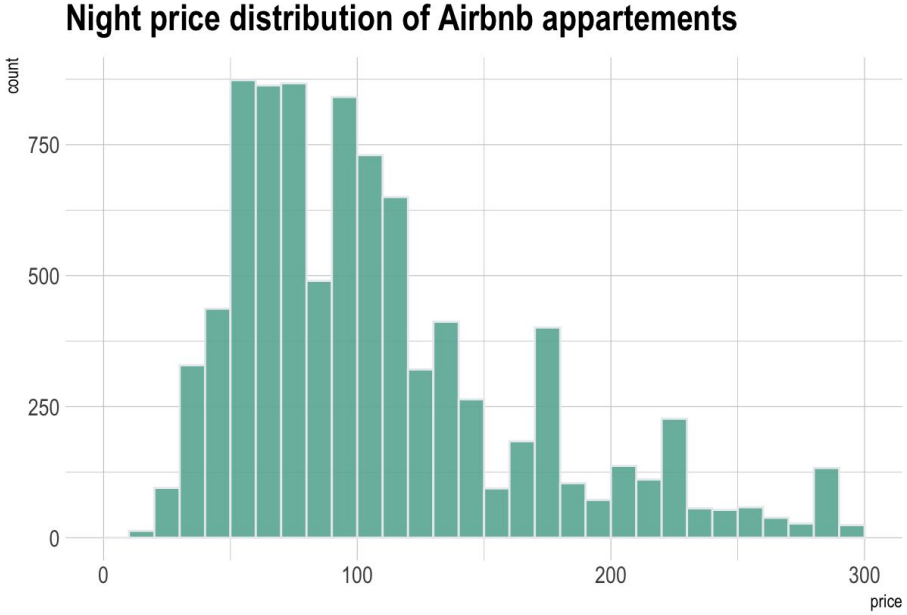
Variables: dimensions and measures.

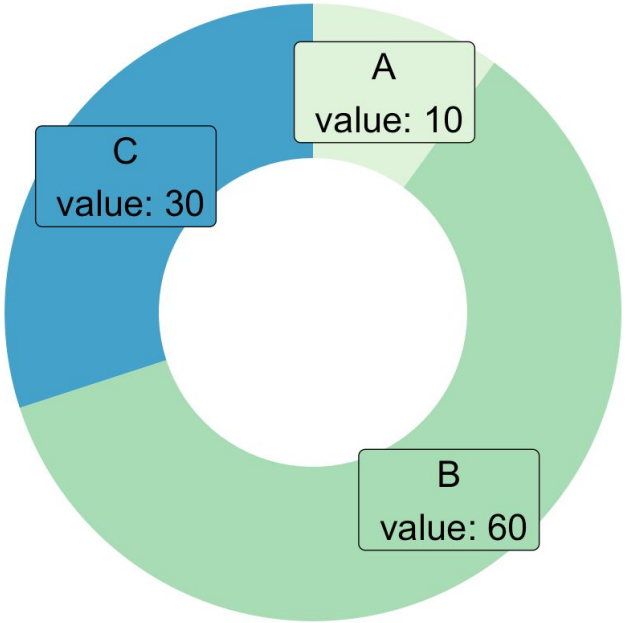
> Patterns: a sampling of options in increasing density.

Group activity: choosing a pattern for a problem.

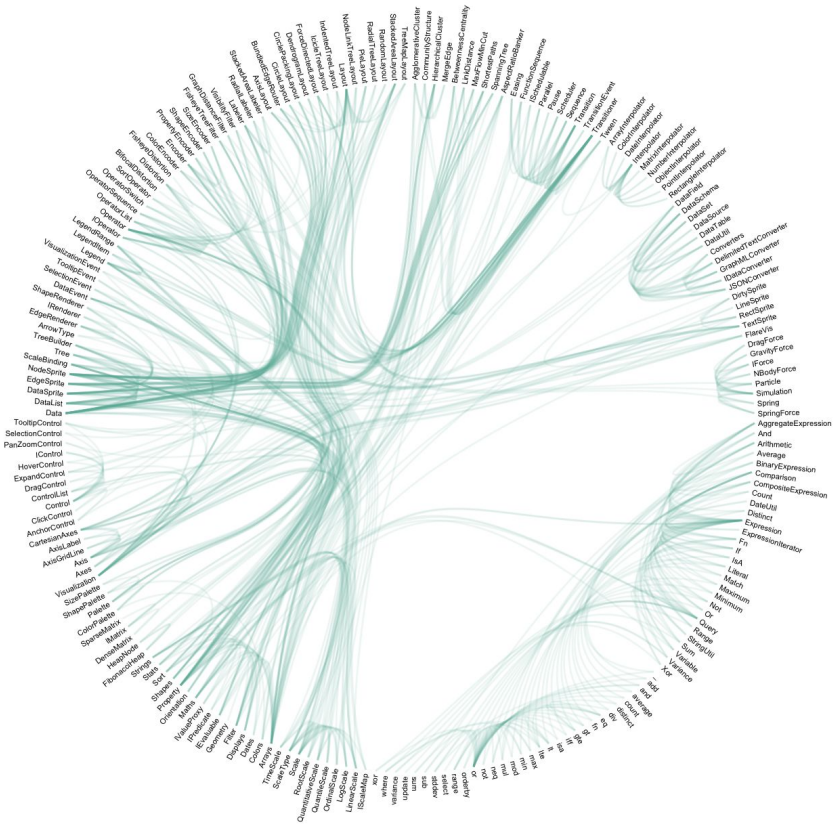
Tasks and domains: a look at contextualizing patterns by understanding users.

1 variable


Night price distribution of Airbnb appartements
count
750
500
250
0
0
100
200
300
price

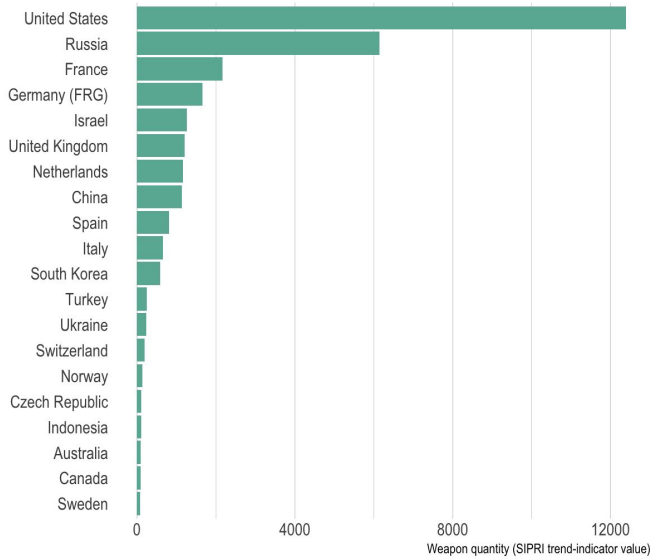

A
value: 10
C
value: 30
B
value: 60

1 variable: hierarchy / graph

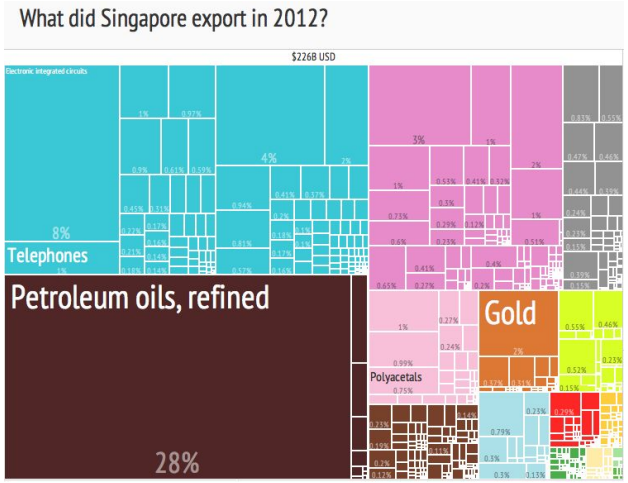


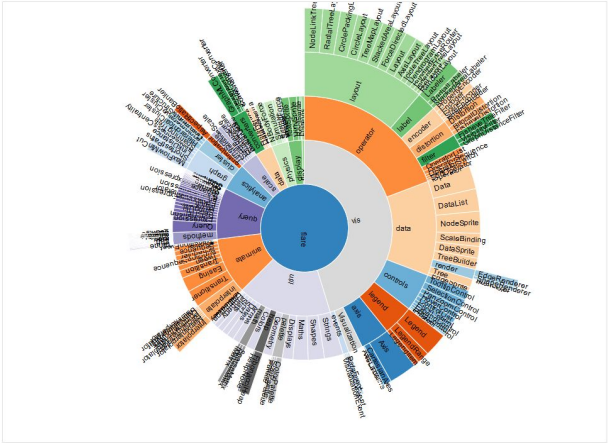
2 variable

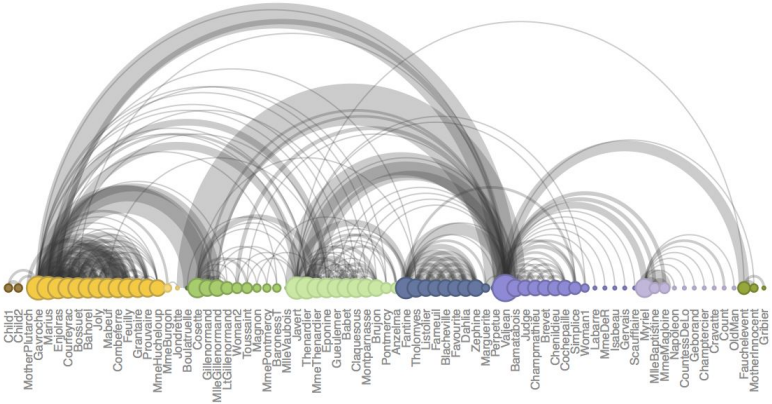

Ground living area partially explains sale price of apartments
Sale price (k$)
0
200
400
600
1000
2000
3000
4000
5000
Ground living area

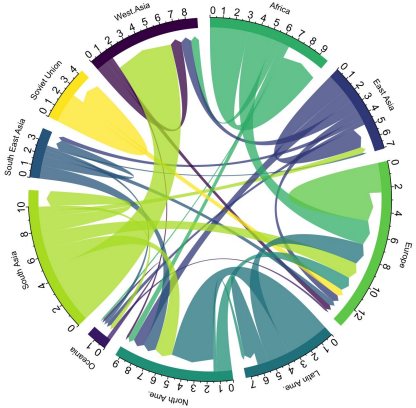

United States
Russia
France
Germany (FRG)
Israel
United Kingdom
Netherlands
China
Spain
Italy
South Korea
Turkey
Ukraine
Switzerland
Norway
Czech Republic
Indonesia
Australia
Canada
Sweden
0
4000
8000
12000
Weapon quantity (SIPRI trend-indicator value)

2 variables: hierarchy / graph

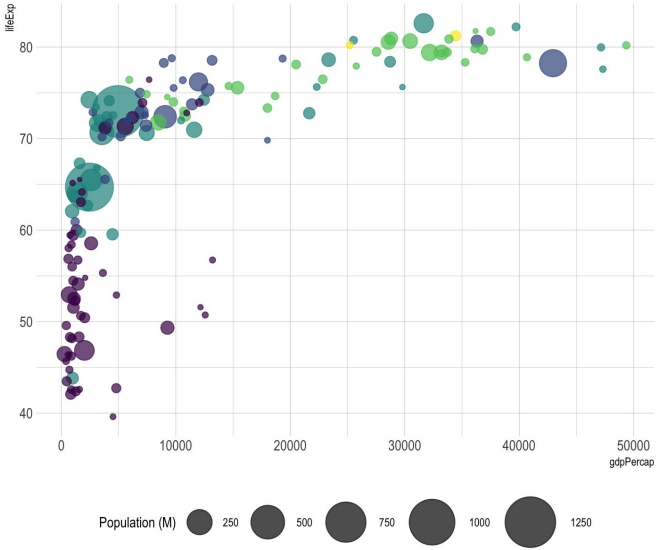


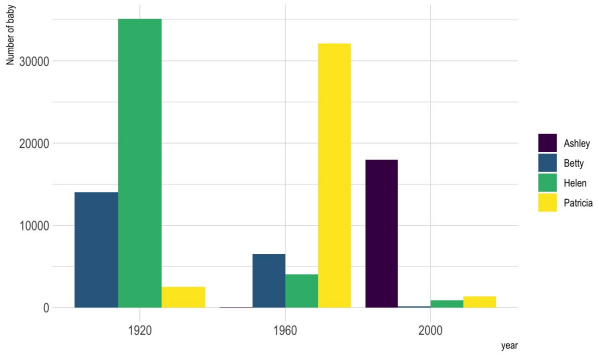


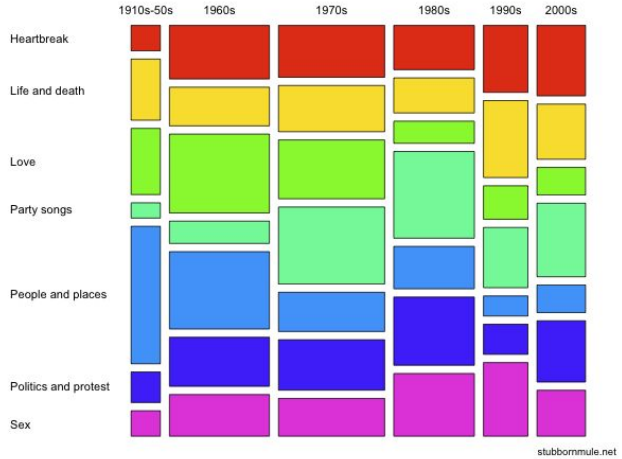




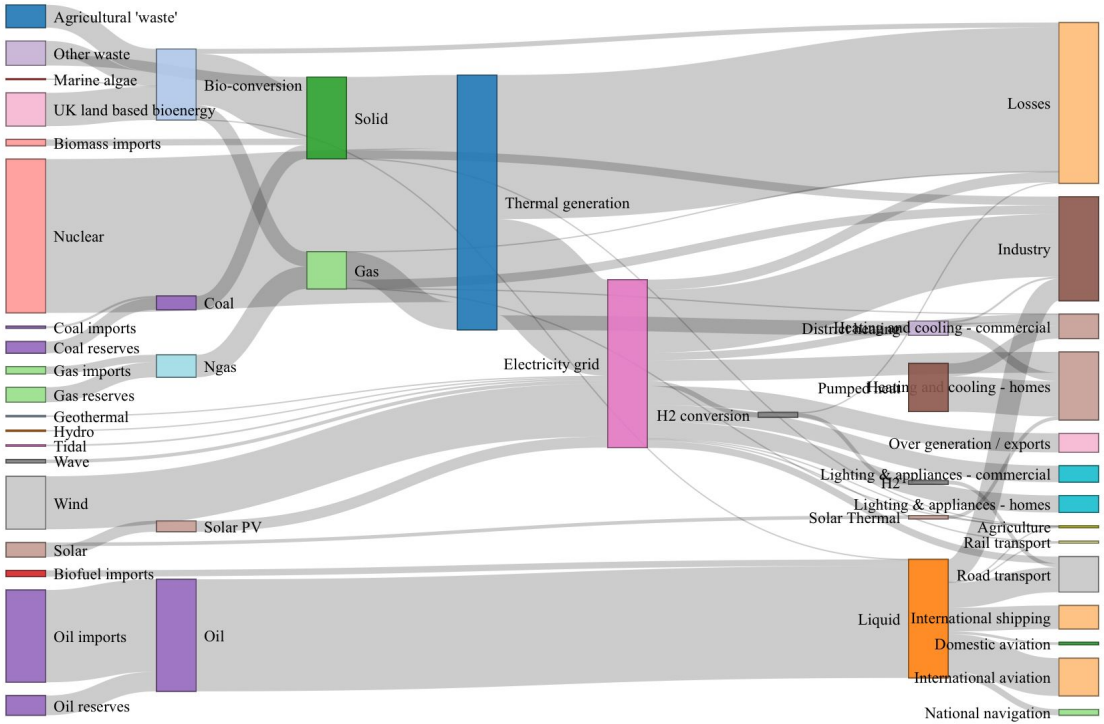
3 variable


lifeExp
80
70
60
50
40
0
10000
20000
30000
40000
50000
gdpPercap
Population (M)
250
500
750
1000
1250

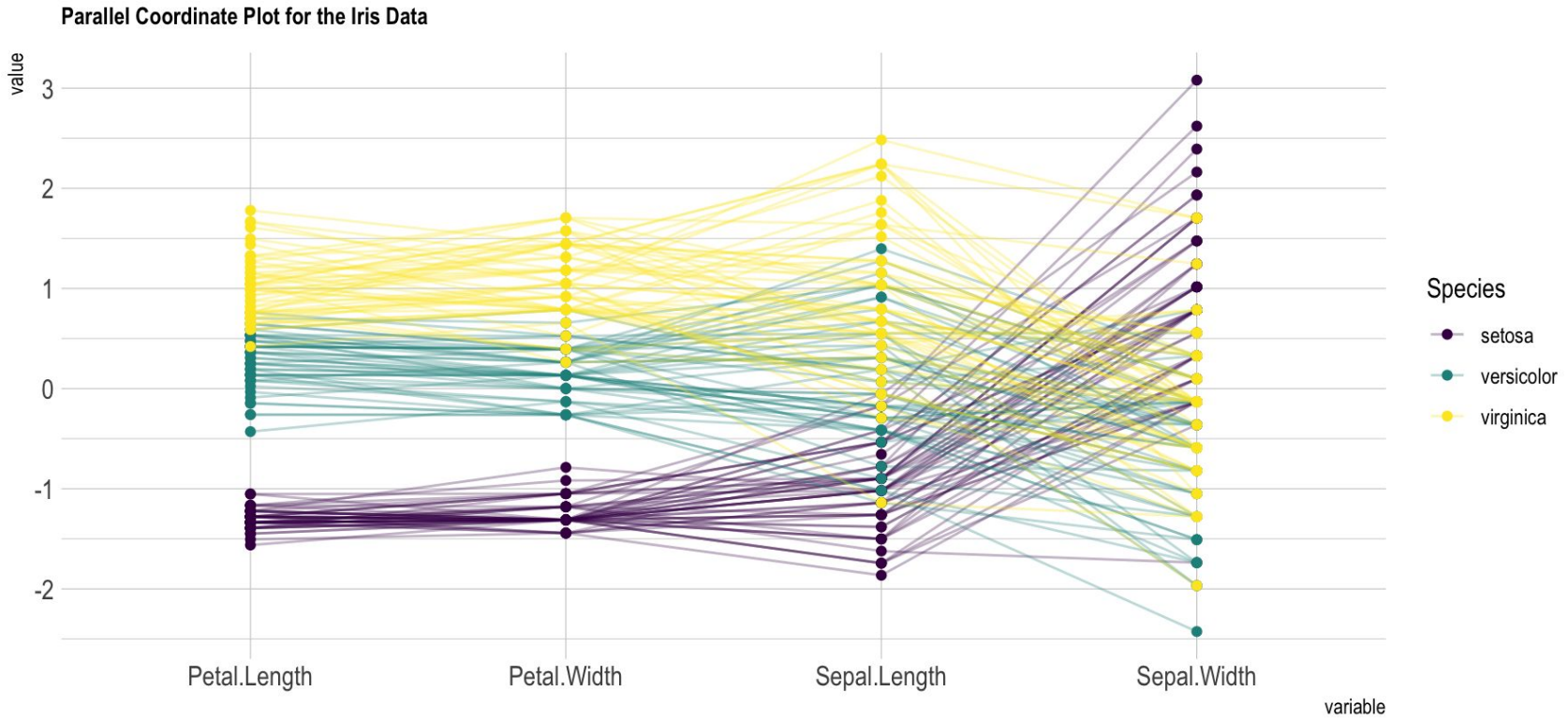

Number of baby
30000
20000
10000
0
1920
1960
2000
year
Ashley
Betty
Helen
Patricia


1910s-50s
1960s
1970s
1980s
1990s
2000s
Heartbreak
Life and death
Love
Party songs
People and places
Politics and protest
Sex
stubbornmule.net

3 variable: hierarchy / graph



4 variable


Parallel Coordinate Plot for the Iris Data
value
3
2
1
0
-1
-2
Petal.Length
Petal.Width
Sepal.Length
Sepal.Width
variable
Species
setosa
versicolor
virginica

The grammar is limited

Especially for higher density plots, the grammar of graphics is limited. We have to think about:

- Shared / dual axes
- Novel representations
- Interactivity / multiple plots

Today

Variables: dimensions and measures.

Patterns: a sampling of options in increasing density.

> Group activity: choosing a pattern for a problem.

Tasks and domains: a look at contextualizing patterns by understanding users.

Example 1

Plot median income by occupation.

Dimension: Occupation
Measure: Income
Representation: Bar

Example 2

Plot median income by occupation and gender.

Example 3

Plot median income and unemployment by occupation and gender.

Example 4

Plot median income, unemployment, and number of people by occupation and gender.

Example 5

Plot median income, unemployment, and number of people by occupation, gender, and race.

Example 6

Plot median income, unemployment, and number of people by occupation, gender, and race in 2015 vs 2025.

Example 7

Plot median income, unemployment, and number of people by occupation, gender, and race each year from 2015 to 2025.

Today

Variables: dimensions and measures.

Patterns: a sampling of options in increasing density.

Group activity: choosing a pattern for a problem.

> Tasks and domains: a look at contextualizing patterns by understanding users.

Re-centering the user

Domains: Who are the users and what are the concepts of the problem area you are working in?

Tasks: What questions is the user trying to answer?

Re-centering the user

Plot median income, unemployment, and number of people by occupation, gender, and race each year from 2015 to 2025.

Domains: researchers - median income, overall unemployment, occupation, gender, race

Tasks: Which group had the largest gender gap by income? Which occupation saw the largest unemployment?

Bring stuff for drawing!

For Monday, please bring pens and paper
to do some drawing by hand.

Citations

Tableau, “Tableau Logo,” Salesforce, 2024. Available: https://www.salesforce.com/news/tableau-from-salesforce-logo-color-1/

H. Wickham, “Tidy Data,” Journal of Statistical Software, 2014. Available: https://www.jstatsoft.org/article/view/v059i10

Tableau, “Dimensions and Measures, Blue and Green,” Salesforce, Available: https://help.tableau.com/current/pro/desktop/en-us/datafields_typesandroles.htm

S. Carmody, “Mosaic Big,” Wikimedia Foundation, 2009. Available: https://en.wikipedia.org/wiki/Mosaic_plot#/media/File:Mosaic-big.png

G. Silvermanaz, “2012 Singapore Products Export Treemap,” Wikimedia Foundation, 2014. Available: https://en.wikipedia.org/wiki/Treemapping#/media/File:2012_Singapore_Products_Export_Treemap.png

S. Maskey, “Zoomable Sunburst with Labels Issue,” StackOverflow, 2014. Available: https://stackoverflow.com/questions/24547620/zoomable-sunburst-with-labels-issue

T. Munzner, "A Nested Model for Visualization Design and Validation," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 921-928, Nov.-Dec. 2009, doi: 10.1109/TVCG.2009.111

S. Zhang, “Canyon,” OpenProcessing. Available: https://openprocessing.org/sketch/2552991

Y. Holtz and C. Healy, “From Data to Viz,” 2018. Available: https://www.data-to-viz.com


CC BY-NC-SA 4.0