CS49000-VIZ - Fall 2020

Introduction to Data Visualization

Task Abstraction

Lecture 7

September 15, 2020

Why?

Actions

Analyze

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*

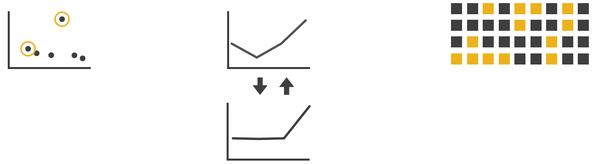
- Produce
  - *Annotate*
  - *Record*
  - *Derive*

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

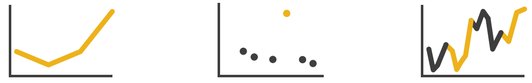
- Identify
- Compare
- Summarize



Targets

All Data

- Trends
- Outliers
- Features

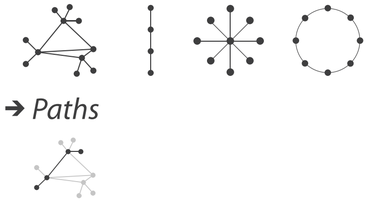


Attributes

- One
- Many
  - *Distribution*
  - *Dependency*
  - *Correlation*
  - *Similarity*
  - *Extremes*

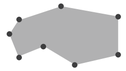
Network Data

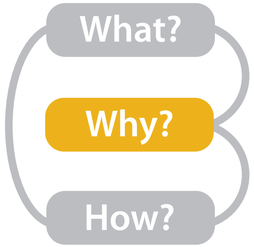
- Topology
- *Paths*



Spatial Data

- Shape





High-level actions: Analyze

- consume
  - –discover vs present
    - aka *explore* vs *explain*
  - –enjoy
    - newcomer
    - aka casual, social

➔ *Discover*

➔ *Present*

➔ *Enjoy*

High-level actions: Analyze

- consume
  - –discover vs present
    - aka *explore* vs *explain*
  - –enjoy
    - newcomer
    - aka casual, social

➔ *Discover*

- Generate new hypotheses
- Check existing hypothesis

➔ *Present*

➔ *Enjoy*

High-level actions: Analyze

- consume
  - –discover vs present
    - aka *explore* vs *explain*
  - –enjoy
    - newcomer
    - aka casual, social

➔ *Discover*

➔ *Present*

- Communicate information
- Story telling
- Guide audience

➔ *Enjoy*

High-level actions: Analyze

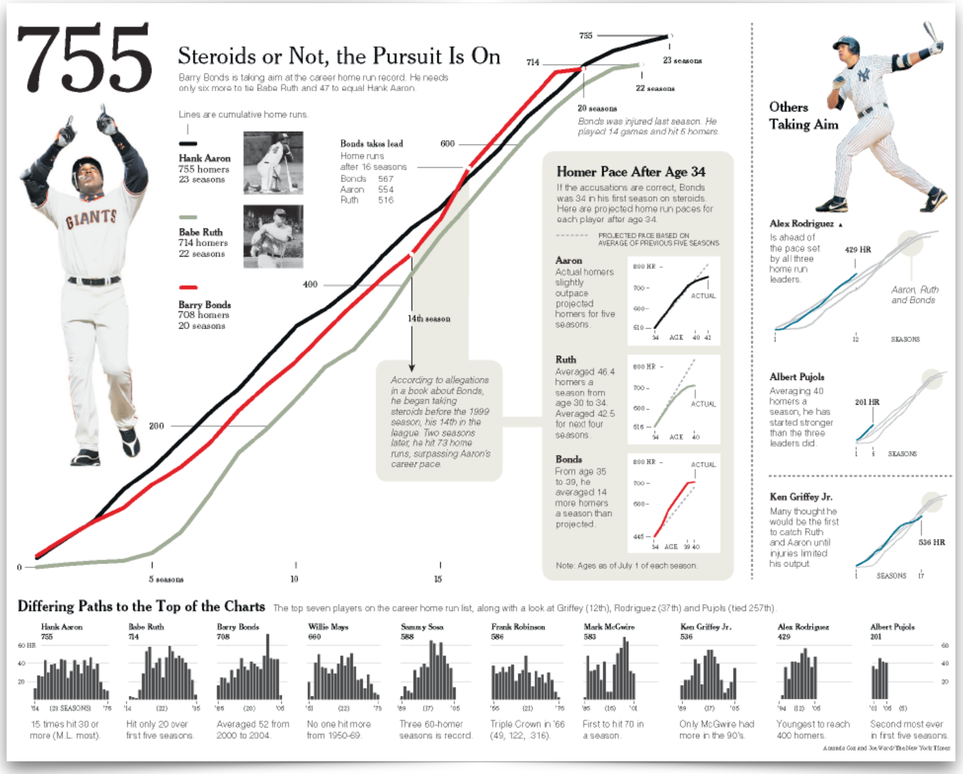
- consume
  - –discover vs present
    - aka *explore* vs *explain*
  - –enjoy
    - newcomer
    - aka casual, social

➔ *Discover*

➔ *Present*

➔ *Enjoy*

Present: Story Telling



Seguel & Heer (credit NYTimes), 2010)

755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons

Babe Ruth
714 homers
22 seasons

Barry Bonds
708 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

755
714
23 seasons
22 seasons
20 seasons

Bonds was injured last season. He played 14 games and hit 5 homers.

600
400
200
0
14th season
5 seasons
10
15

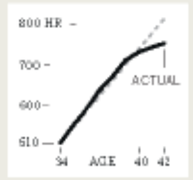
According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

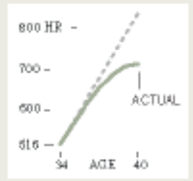
If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

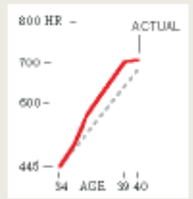
Aaron
Actual homers slightly outpace projected homers for five seasons.



Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.

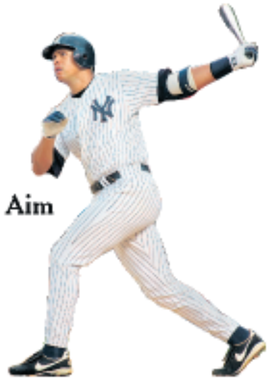


Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

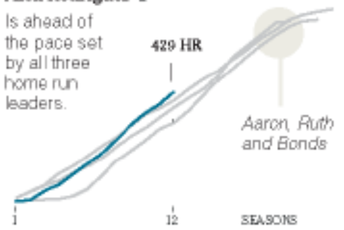


Note: Ages as of July 1 of each season.

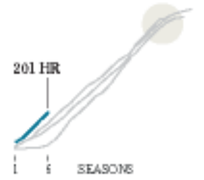
Others Taking Aim



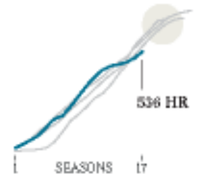
Alex Rodriguez ▲
Is ahead of the pace set by all three home run leaders.



Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

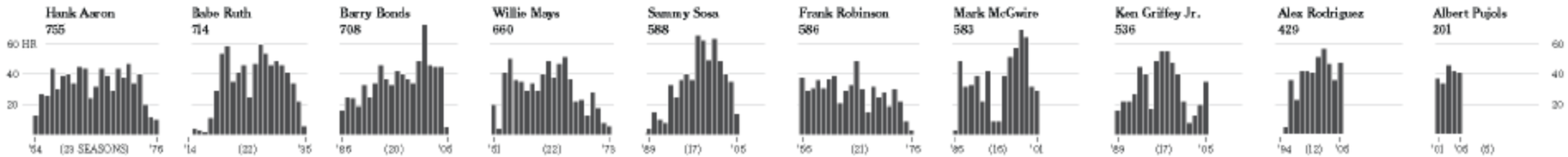


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.



Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).



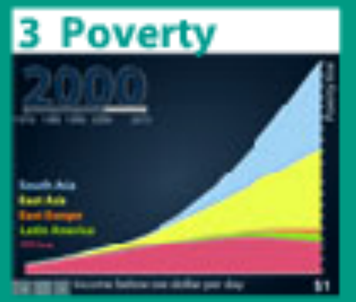
Amanda Cox and Joe Ward/The New York Times

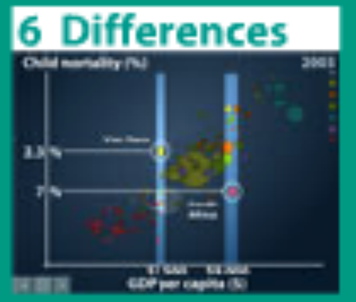
Human Development Trends 2005

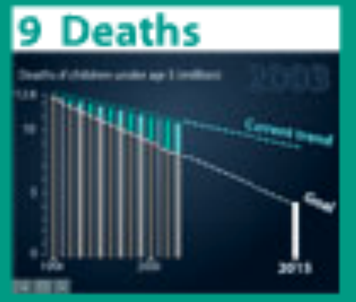
Interactive presentation of some of the messages in the Human Development Report 2005

English
Dansk
Portuguese
Suomi
Français
Deutsch

Produced in collaboration with:
GAPMINDER
www.gapminder.org





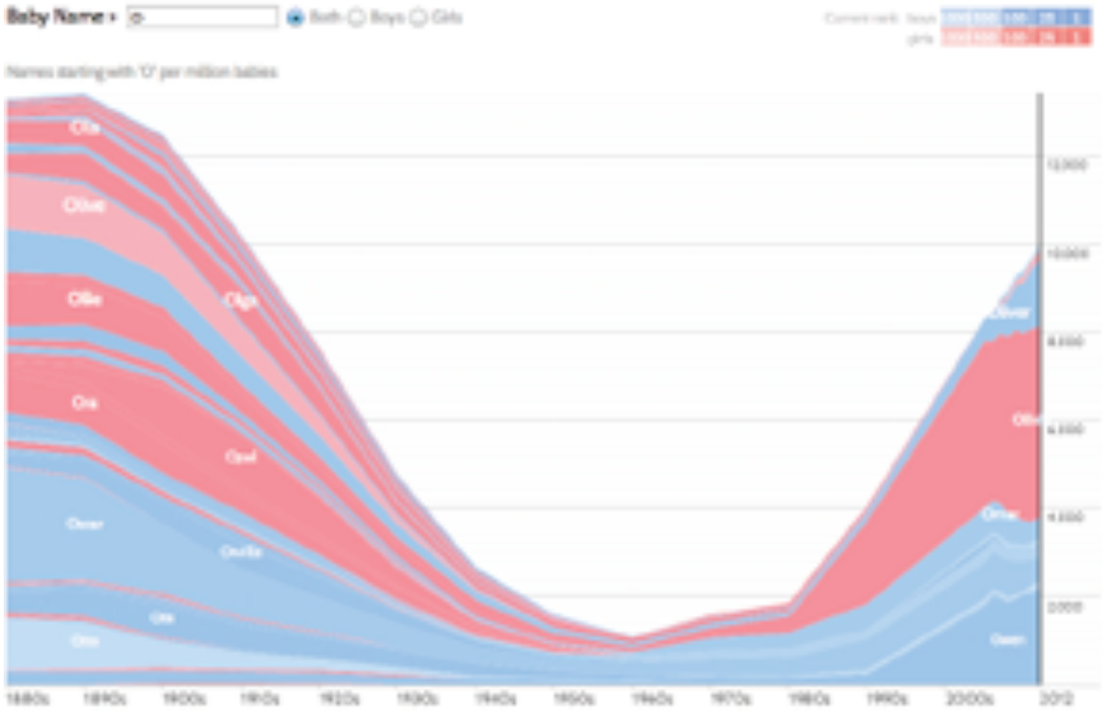


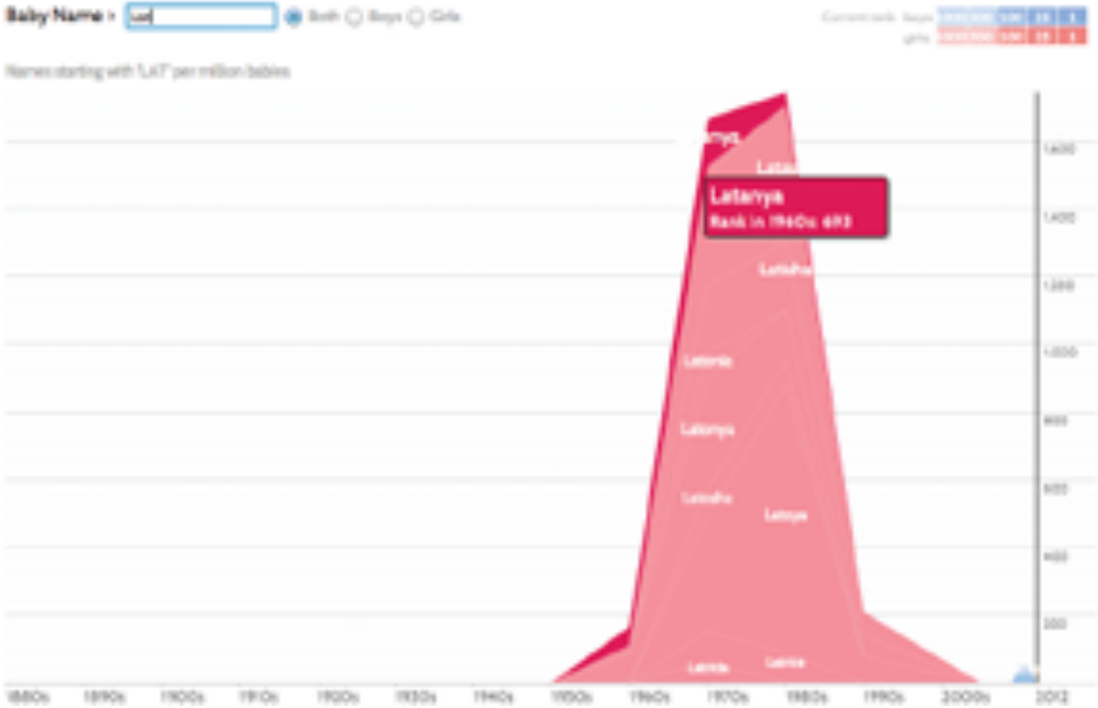
English translation: Claes Johansson, UNDP

1 2 3 4 5 6 7 8 9

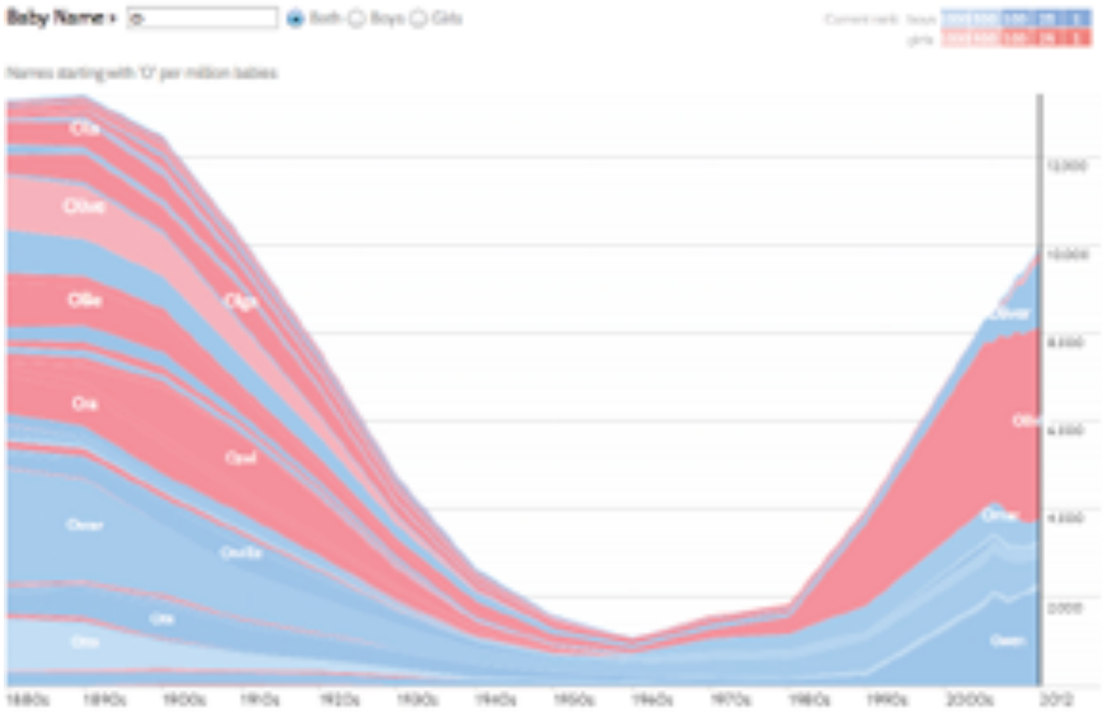
Start

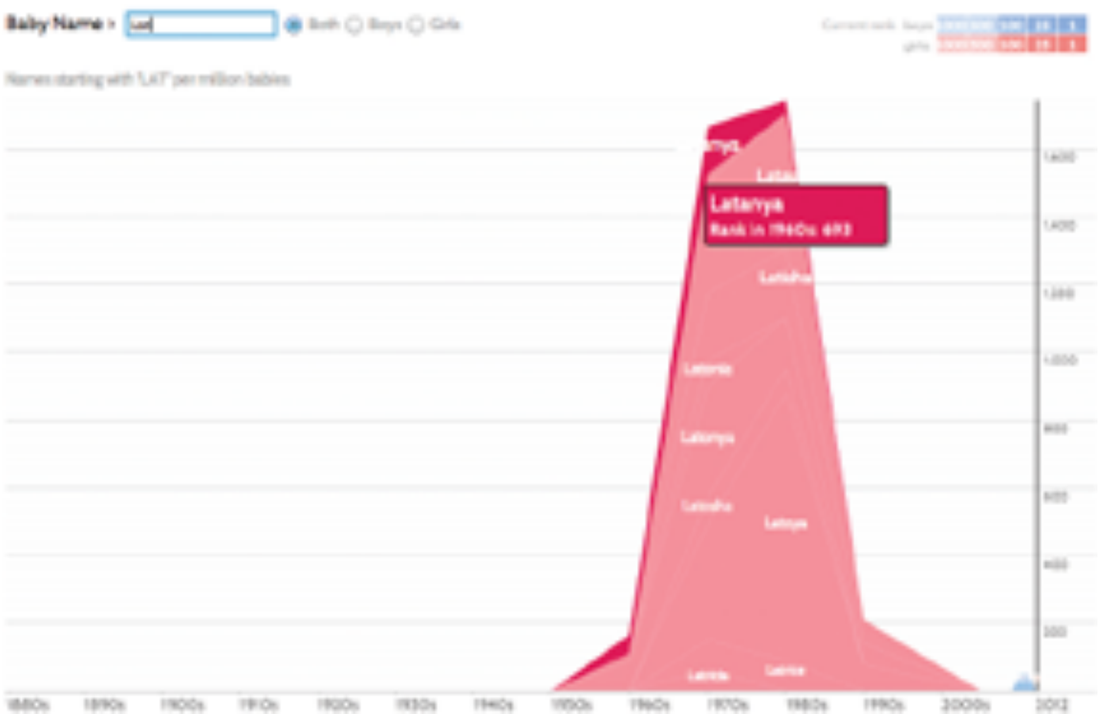
BabyNameWizzard





BabyNameWizzard

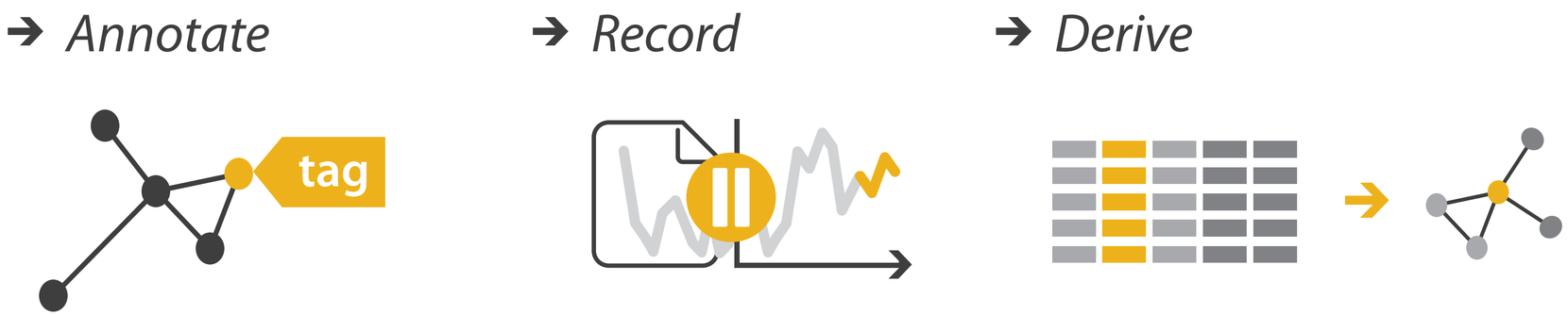




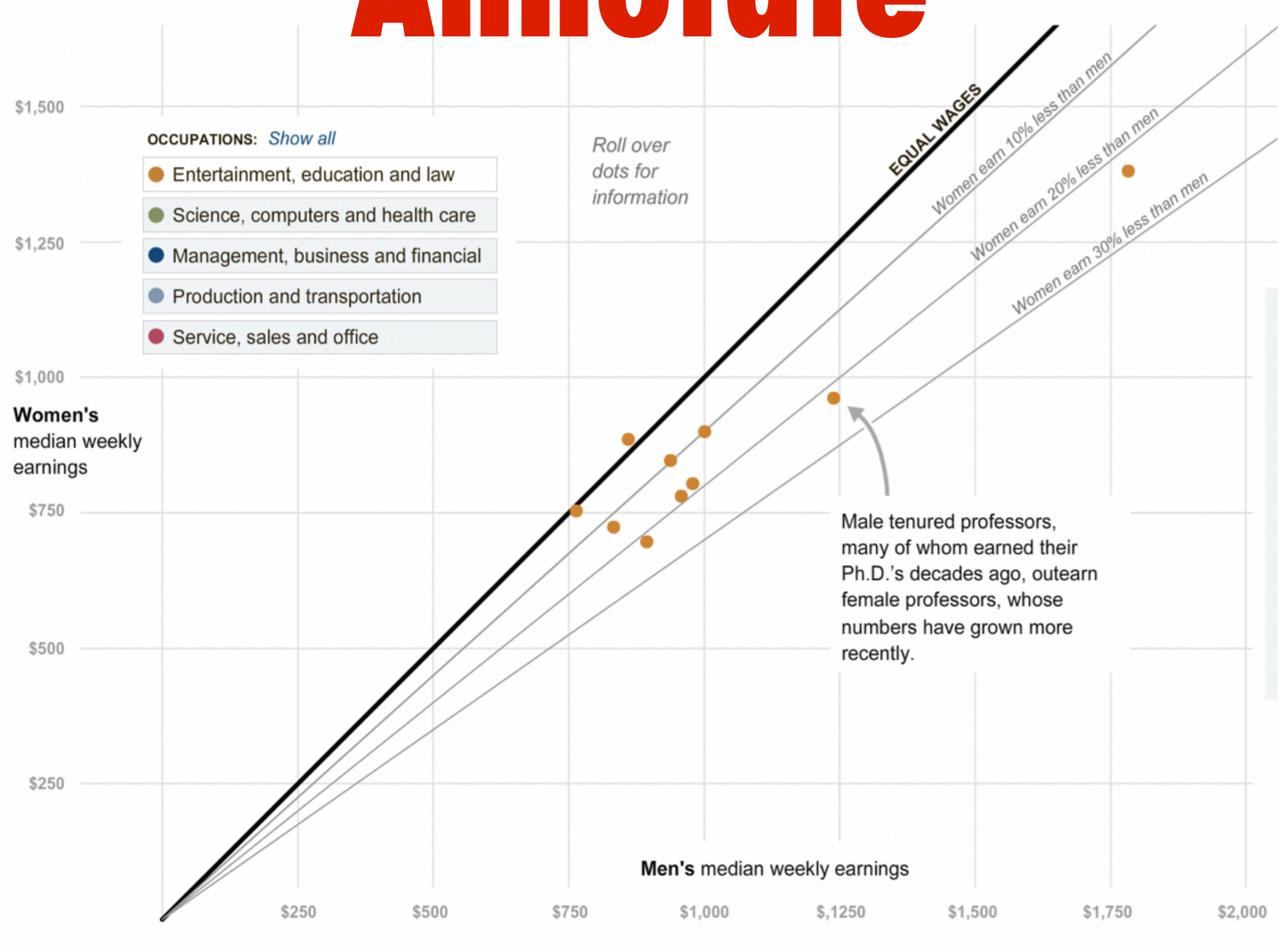
http://www.babynamewizard.com/voyager

High-level actions: Analyze

- produce
  - –annotate, record
  - –derive
    - crucial design choice



Annotate

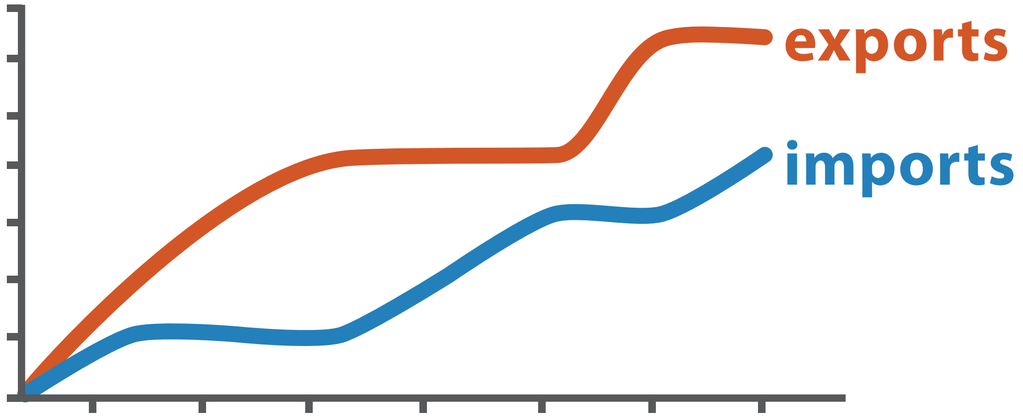

$1,500
$1,250
$1,000
$750
$500
$250
OCCUPATIONS: Show all
Entertainment, education and law
Science, computers and health care
Management, business and financial
Production and transportation
Service, sales and office
Roll over dots for information
EQUAL WAGES
Women earn 10% less than men
Women earn 20% less than men
Women earn 30% less than men
Women's median weekly earnings
Male tenured professors, many of whom earned their Ph.D.'s decades ago, outearn female professors, whose numbers have grown more recently.
Men's median weekly earnings
$250
$500
$750
$1,000
$1,250
$1,500
$1,750
$2,000

Derive

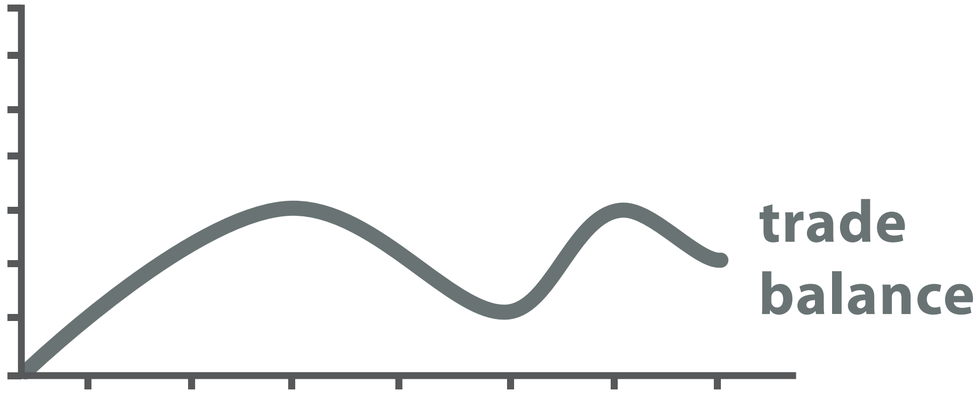
- don’t just draw what you’re given!
  - decide what the right thing to show is
  - create it with a series of transformations from the original dataset
  - draw that

Derive

- don’t just draw what you’re given!



Original Data



trade balance = *exports* – *imports*

Derived Data

Derive

- Don’t just draw what you’re given!
- One of the four major strategies for handling complexity

Actions: Mid-level search low-level query

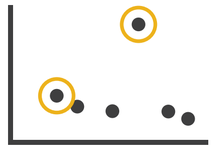
- What does the user know?
  - Target, location

	Target known	Target unknown
Location known	 *Lookup*	 *Browse*
Location unknown	 *Locate*	 *Explore*

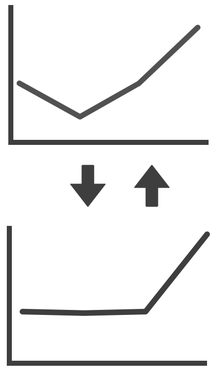
Actions: low-level query

- How much of the data matters?
  - One, some, all

→ Identify

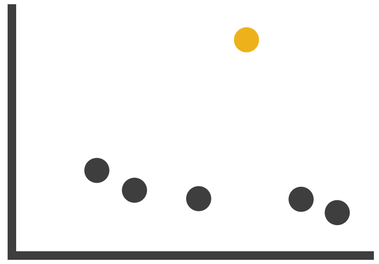
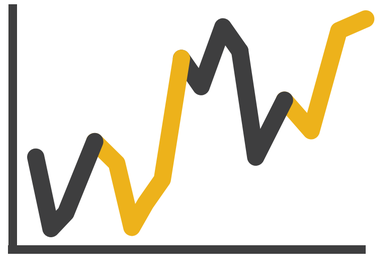


→ Compare



→ Summarize

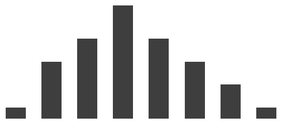
Why: Targets

- All Data
  - Trends
    
  - Outliers
    
  - Features
    

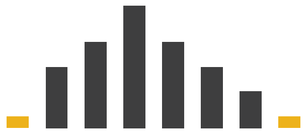
Why: Targets

Attributes

- One
  - *Distribution*

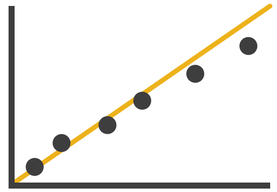


  - *Extremes*

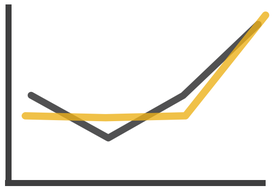


- Many
  - *Dependency*

  - *Correlation*



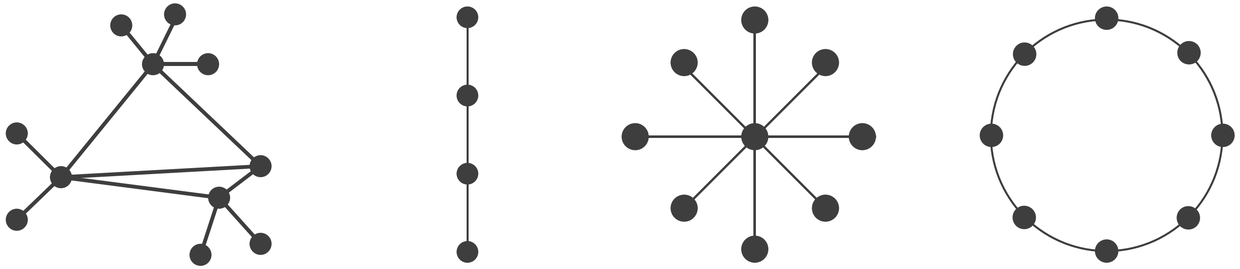
  - *Similarity*



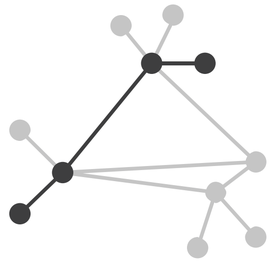
Why: Targets

➔ Network Data

➔ Topology



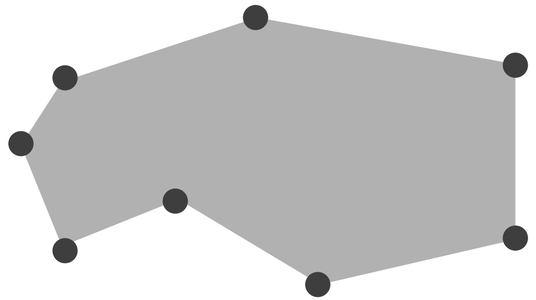
➔ *Paths*



Why: Targets

→ Spatial Data

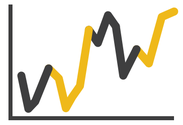
→ Shape



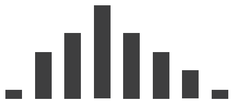
Why: Targets

- ALL DATA
  - Trends
  - Outliers
  - Features





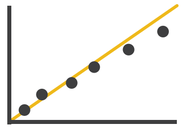
- ATTRIBUTES
  - One
    - *Distribution*



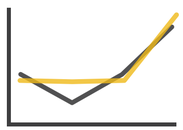
    - *Extremes*

  - Many
    - *Dependency*

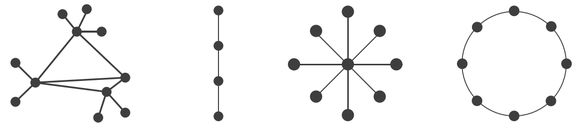
    - *Correlation*



    - *Similarity*

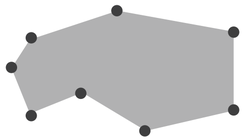


- NETWORK DATA
  - Topology



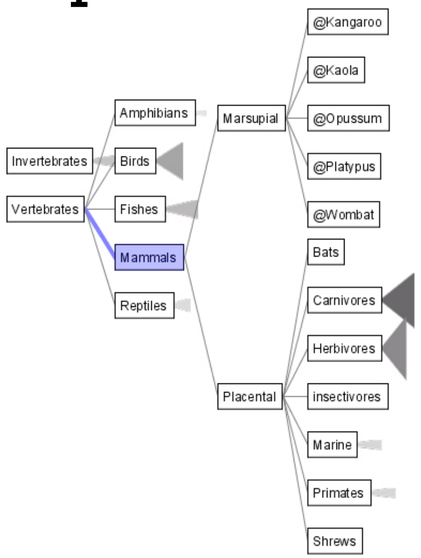
  - *Paths*

- SPATIAL DATA
  - Shape

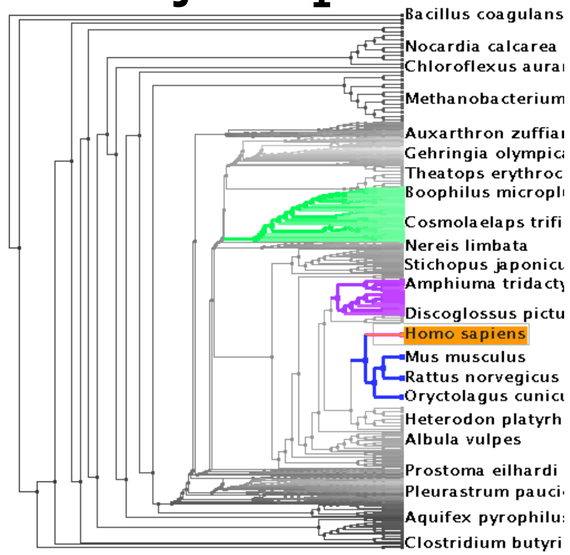


Analysis example: Compare idioms

SpaceTree

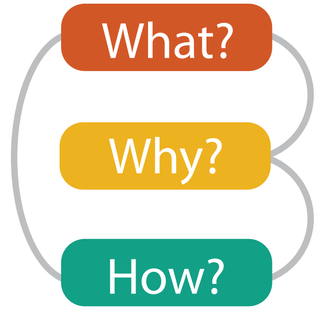


TreeJuxtaposer



[SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Grosjean, Plaisant, and Bederson. Proc. InfoVis 2002, p 57–64.]

[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453– 462, 2003.]



What?

- Tree

Why?

- Actions
  - Present
  - Locate
  - Identify

- Targets
  - Path between two nodes

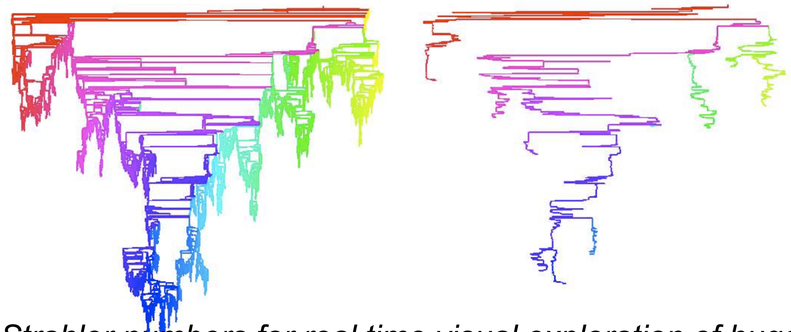
How?

- SpaceTree
  - Encode
  - Navigate
  - Select
  - Filter
  - Aggregate

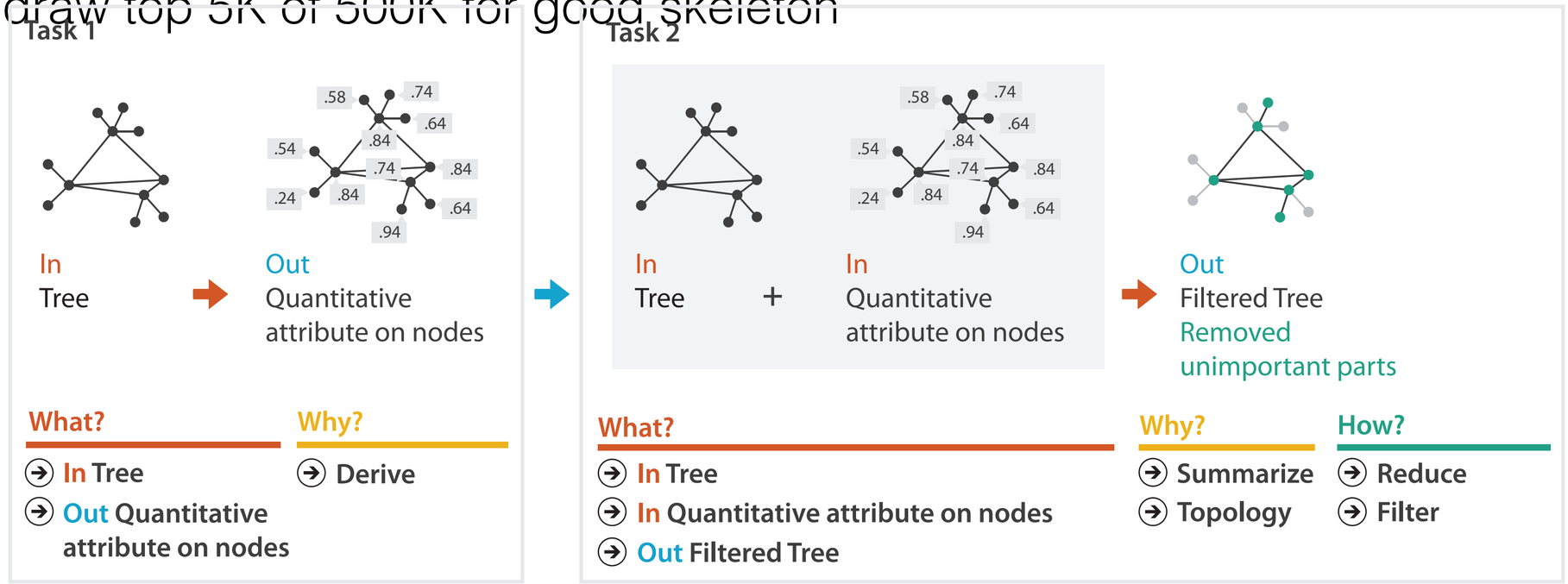
- TreeJuxtaposer
  - Encode
  - Navigate
  - Select
  - Arrange

Analysis example: Derive one attribute

- Strahler number
  - centrality metric for trees/networks
  - derived quantitative attribute
  - draw top 5K of 500K for good skeleton



[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]



Chained sequences

- Output of one is input to next
  - express dependencies
  - separate means from ends