Why does visualization work?
- limits of memory & cognition
- power of perception to reveal
  - how many V's?

Exercise
- Which gender and income level shows a different effect of age on triglyceride levels?

Why analyze visualizations?
- imposes structure on huge design space
  - scaffold to help you think systematically about choices
- analyzing existing as stepping stone to designing new
- most possibilities ineffective for particular tasks/combination

Abstractions: Nested Model

Exercise
- Which gender and income level shows a different effect of age on triglyceride levels?

Why does visualization work?
- limits of memory & cognition
- power of perception to reveal
  - how many V's?

Exercise
- Which gender and income level shows a different effect of age on triglyceride levels?

Different threats to validity at each level
- cascading effects downstream

Mismatches: Common problem

Exercise
- Which gender and income level shows a different effect of age on triglyceride levels?

Why analyze visualizations?
- imposes structure on huge design space
  - scaffold to help you think systematically about choices
- analyzing existing as stepping stone to designing new
- most possibilities ineffective for particular tasks/combination

Abstractions: Nested Model

Exercise
- Which gender and income level shows a different effect of age on triglyceride levels?

Why does visualization work?
- limits of memory & cognition
- power of perception to reveal
  - how many V's?

Exercise
- Which gender and income level shows a different effect of age on triglyceride levels?
**Dataset Types**

- **Networks/Graphs**
  - nodes (vertices) connected by links (edges)
  - tree is special case: no cycles
  - often have roots and are directed

- **Spatial Fields**
  - attribute values associated with cells
  - cell contains value from continuous domain
  - measured or simulated

- **Spatial Properties (Spatial) Fields**
  - eg temperature, pressure, wind velocity
  - measured or simulated

- **Multidimensional Tables**
  - one item per row
  - each column is attribute
  - cell holds value
  - indexing based on multiple keys
  - eg genes, patients

- **Trees**
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair
  - unique key (could be implicit)

**Other Data Types**

- **Spatial Data**
  - attribute values associated with cells
  - cell contains value from continuous domain
  - measured or simulated

- **Geometry (Spatial)**
  - attribute values associated with cells
  - cell contains value from continuous domain
  - measured or simulated

- **Items**
  - individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"

- **Attributes**
  - name, age, shirt size, fave fruit
  - "dependent variable"

**Items & Attributes**

- **Items**
  - individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"

- **Attributes**
  - name, age, shirt size, fave fruit
  - "dependent variable"

**What does data mean?**

- **Data** is a collection of information or measurements.

- **Data types** structural or mathematical interpretation of data

- **Data types**
  - item, link, attribute, position, grid

- **Data** can be real-world meaning or semantics:
  - eg temperature, pressure, wind velocity

**Rasters**

- **Raster table**
  - indexing based on multiple keys
  - eg genes, patients

**Networks**

- **Network**
  - nodes (vertices) connected by links (edges)
  - tree is special case: no cycles

**Table**

- **Table**
  - one item per row
  - each column is attribute
  - cell holds value
  - indexing based on multiple keys
  - eg genes, patients
**Spatial fields**
- attribute values associated with cells
- cell contains value from continuous domain
  - eg temperatura, pressure, wind velocity
- measured or simulated
- beyond the scope of this class
  - sampling
  - interpolation
- how to model attributes elsewhere
- grid types

**Spatial fields**
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated
- beyond the scope of this class
  - sampling
  - interpolation
- how to model attributes elsewhere
- grid types

**Dataset types**
- Tables
- Attributes (columns)
- Items 
  (rows)
- Cell containing value
- Networks
- Link
- Node 
  (item)
- Items 
  (nodes)
- Links
- Attributes
- Grids
- Positions
- Attributes
- Items
- Positions
- Items

**Dataset and data types**
- **Dataset Types**
  - Tables
  - Networks & Trees
  - Spatial
  - Fields
  - Geometry
- **Collections**
  - how we group items
  - sets
  - unique items, unordered
  - lists
  - ordered, duplicates possible
  - clusters
  - groups of similar items
- **Geometry**
  - shape of items
  - explicit spatial positions
  - points, lines, curves, surfaces, regions
  - boundary between computer graphics and visualization
  - graphics geometry taken as given
  - via geometry is result of a design decision
- **Attribute types**
  - which classes of values & measurements!
  - categorical (nominal)
    - compare equality
    - no implicit ordering
    - ordered
      - ordinal
      - less/greater than defined
      - quantitative
    - meaningful magnitude
    - arithmetic possible
  - numerical
    - positive
    - negative
    - magnitude
    - ratio (e.g., 5:1)
    - power (e.g., 10^2)
    - time (e.g., 10 seconds)
  - string
    - text
    - characters
    - sequence (e.g., DNA)

**Hierarchical data**
- multi-level structure
  - space
  - time
  - others
- example: zipdecode

**Data vs conceptual models**
- data model
  - mathematical abstraction
  - sets with operations, eg floats with +, /, *
  - variable data types in programming/languages
  - conceptual model
    - mental construction (semantics)
    - supports reasoning
    - typically based on understanding of tasks [stay tuned, next week]
  - data abstraction process relies on conceptual model
    - for transforming data if needed

**Data vs conceptual models, example**
- data model floats
  - 1232.546, -14.35
  - conceptual model
    - temperature
    - multiple possible data abstractions
      - continuous to 2 significant figures: quantitative
        - task: forecasting the weather
        - hot, warm, cold: ordinal
        - task: deciding if bath water is ready
      - above freezing, below freezing: categorical
    - data abstraction process relies on conceptual model
      - for transforming data if needed

**Data abstraction, three operations**
- translate from domain-specific language to generic visualization language
- identify dataset type(s), attribute types
- identify cardinality
  - how many items in the dataset?
  - what is cardinality of each attribute?
- range for quantitative data
- consider whether to transform data
  - guided by understanding of task

**Data vs conceptual model, example**
- data model floats
  - 1232.546, -14.35
  - conceptual model
    - temperature
    - multiple possible data abstractions
      - continuous to 2 significant figures: quantitative
        - task: forecasting the weather
        - hot, warm, cold: ordinal
        - task: deciding if bath water is ready
      - above freezing, below freezing: categorical
        - task: deciding if I should leave the house today
### Data and Dataset Types

**Grids**
- Diverging
- Sequential
- Ordered
- Nested

**Networks**
- Link
- Node

**Geometry**
- Derived attributes: compute from originals
- simple change of type
- acquire additional data
- complex transformation
- more on this next time

### Data abstraction practice

**2018 Central Park Squirrel Census**

[https://www.thecitycrux.com/case/](https://www.thecitycrux.com/case/)

### Abstract Tasks (Why)

- **Domain characterization**
  - details of an application domain
  - group of users, target domain, their questions, & their data
  - varies widely by domain
  - must be specific enough to get traction
  - domain questions/problems
  - break down into simpler abstract tasks

- **Design Process**
  - **Characterize Domain Situation**
  - **Map Domain-Language to Data Abstraction**
  - **Map Domain-Language Task to Abstract Task**
  - **Identify/Create Suitable Idiom/Technique**
  - **Identify/Create Suitable Algorithm**

### Example: Find good movies
- **domain**
  - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Data abstraction practice**
  - 2018 Central Park Squirrel Census

- **Actions/Analyze**
  - actions
    - analyze
    - high-level choices
    - search
    - find a known/unknown item
    - query
  - produce
    - annotate, record
    - derive
    - crucial design choice

- **Analysis example: Derive one attribute**
  - [Stringer number]
  - centrality metric for trees/networks
  - derived quantitative attribute
  - draw top 50 of 1000 for good evaluation

- **Why?**
  - high-level abstraction

- **How?**
  - multiple abstraction layers
  - details of an application domain
  - abstract layer

### Example: Find good movies
- **How?**
  - multiple abstraction layers
  - details of an application domain
  - abstract layer

- **Why?**
  - high-level abstraction

- **Analysis example: Derive one attribute**
  - [Stringer number]
  - centrality metric for trees/networks
  - derived quantitative attribute
  - draw top 50 of 1000 for good evaluation

### Example: Find good movies
- **Task abstraction: Actions and targets**
  - very high-level pattern
  - (category, target) pairs
  - (category, features)
  - consume
  - produce
  - analyze
  - search
  - find a known/unknown item
  - query
  - produce
  - annotate, record
  - derive
  - crucial design choice

- **Actions: Analyze**
  - consume
  - produce
  - derive
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation

### Example: Find good movies
- **Example: Find good movies**
  - **domain**
    - general population, movie enthusiasts

- **Task 1**
  - identify good movies in genres I like
  - domain:
    - general population, movie enthusiasts
    - slightly different data

- **Derived abstract task**
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- **Algorithm**
  - efficient computation
Example: Economics
• task: compare and derive
• data: derive change
69
The Economist

Example: Task abstraction in genomics
You have been approached by a geneticist to help with a visualization problem. She has gene expression data (data that measures the activity of the genes) for 30 cancer tissue samples. She is applying an experimental drug to see whether the cancer tissue dies as she hopes, but she finds that only some samples show the desired effect. She believes that the difference between the samples is caused by differential expression (different activity) of genes in a particular pathway, i.e., an interaction network of genes. She would like to understand which genes are likely to cause the difference, and what role they play in that pathway.

Example: Task abstraction in genomics
• ... only some samples show the desired effect
• ... the difference between the samples is caused by differential expression (different activity) of genes in a particular pathway. She would like to understand which genes are likely to cause the difference

Example: Task abstraction in genomics
• ... which genes are likely to cause the difference, and what role they play in that pathway.

Example: Task abstraction in genomics
• ... locate the outlier in the network

Example: Task abstraction in genomics
• ... explore the topology

Credits
• Visualization Analysis and Design (Ch 1, 2, 3, 4)
• Alex Lex & Miriah Meyer: http://dataiscourse.net/

Assignments
• Programming 0
• Foundations 1

Example: Economics
• task: compare and derive
• data: derive change
69
The Economist

Example: Task abstraction in genomics
You have been approached by a geneticist to help with a visualization problem. She has gene expression data (data that measures the activity of the genes) for 30 cancer tissue samples. She is applying an experimental drug to see whether the cancer tissue dies as she hopes, but she finds that only some samples show the desired effect. She believes that the difference between the samples is caused by differential expression (different activity) of genes in a particular pathway, i.e., an interaction network of genes. She would like to understand which genes are likely to cause the difference, and what role they play in that pathway.

Example: Task abstraction in genomics
• ... only some samples show the desired effect
• ... the difference between the samples is caused by differential expression (different activity) of genes in a particular pathway. She would like to understand which genes are likely to cause the difference

Example: Task abstraction in genomics
• ... which genes are likely to cause the difference, and what role they play in that pathway.

Example: Task abstraction in genomics
• ... locate the outlier in the network

Example: Task abstraction in genomics
• ... explore the topology

Credits
• Visualization Analysis and Design (Ch 1, 2, 3, 4)
• Alex Lex & Miriah Meyer: http://dataiscourse.net/

Assignments
• Programming 0
• Foundations 1