

Data Visualization II

DSCI 532

Lecture 1 - January 2, 2019

Cydney Nielsen

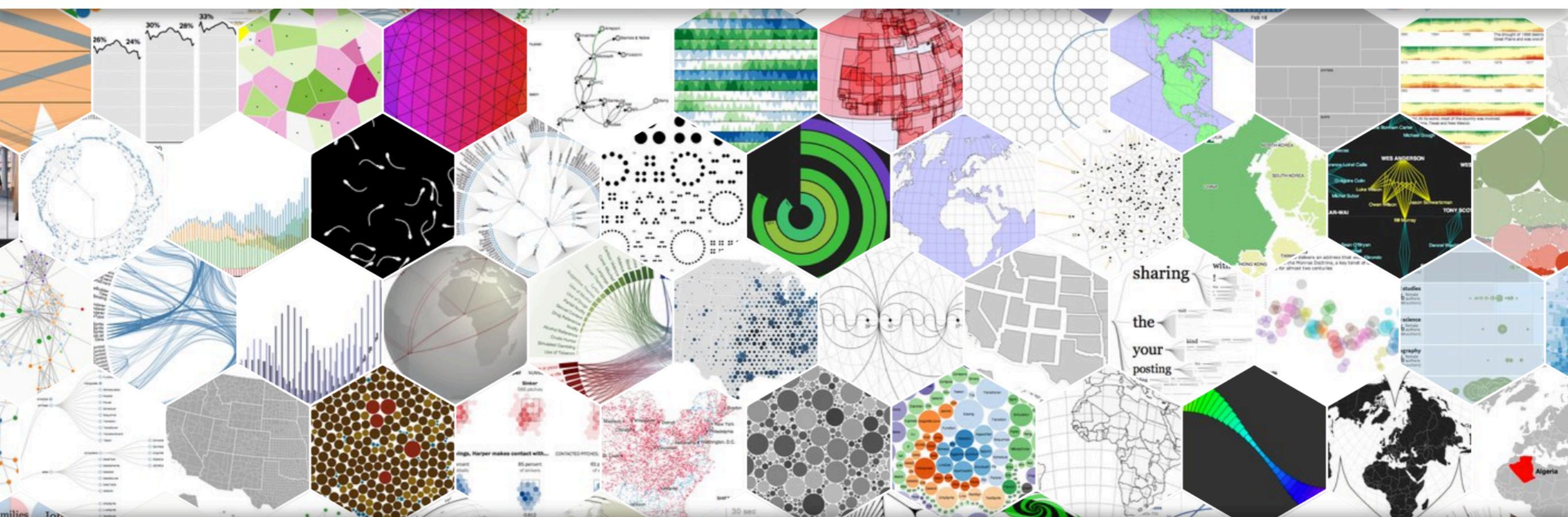
Senior Designer, Microsoft
Adjunct Professor, UBC Department of Computer Science

Why are we here?

Value of Visualization Skills

- **Make us better data communicators**
 - Ability to take complex information and make it interpretable
- **Make us savvy data consumers**
 - Ability to think critically about data representation

Course Scope



2D graphics on desktop display

<https://d3js.org>

Course Scope



<https://ind.pn/2s1HZWI>

NOT 3D graphics | AR | VR

Course Structure

- **Lectures**
 - Equip you with principles to approach building effective ***interactive*** data visualizations
 - Practice those principles through design critiques
- **Labs**
 - Hands-on with shiny
 - Opportunity to build your own shiny app
 - Apply the principles and approaches from lecture

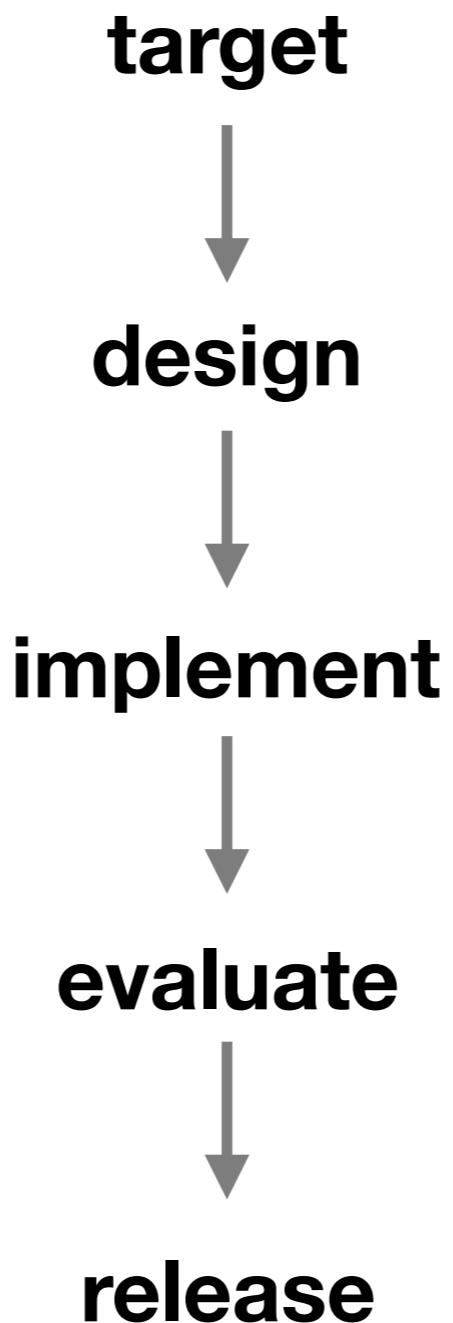
Details

- **Office Hours**
 - 8:30-9:00 am Mon/Wed before class
- **Questions / Comments**
 - UBC MDS Slack channel for DSCI 532
- **Slides available on GitHub**

Process

How to design a data visualization

Process



Target

target



design



implement



evaluate



release

What is the purpose of my visualization?

- What am I trying to show?
- What questions am I trying to answer?

Who is my audience?

- Is it for me?
- Is it for executive decision makers?
- Is it for data analysts?
- Is it for the general public?

Real world example

Genome assembly visualization at BC Cancer

“Here’s my data, how should I visualize it?”

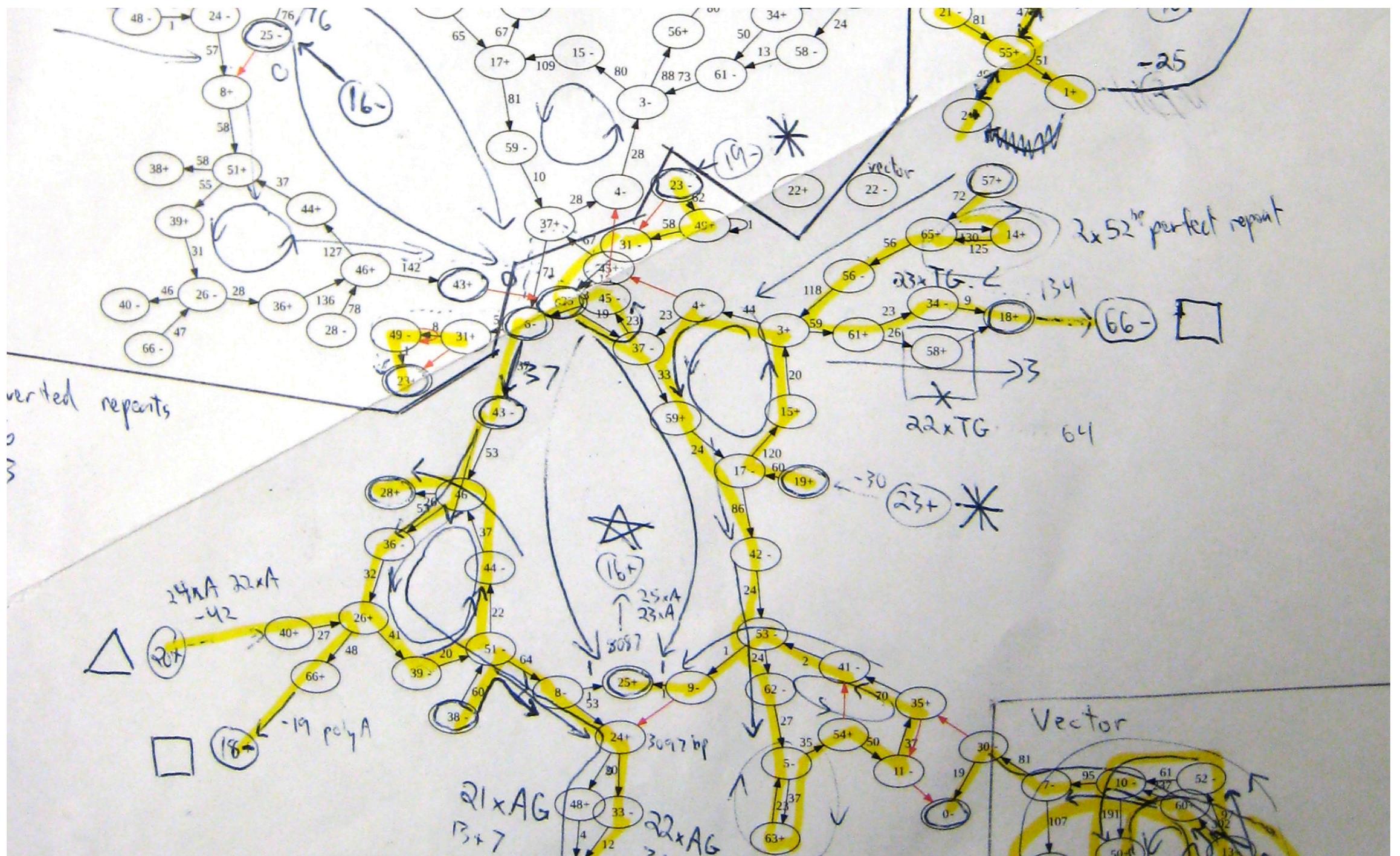


Image credit: Shaun Jackman, BC Cancer

"Here's my data, how should I visualize it?"

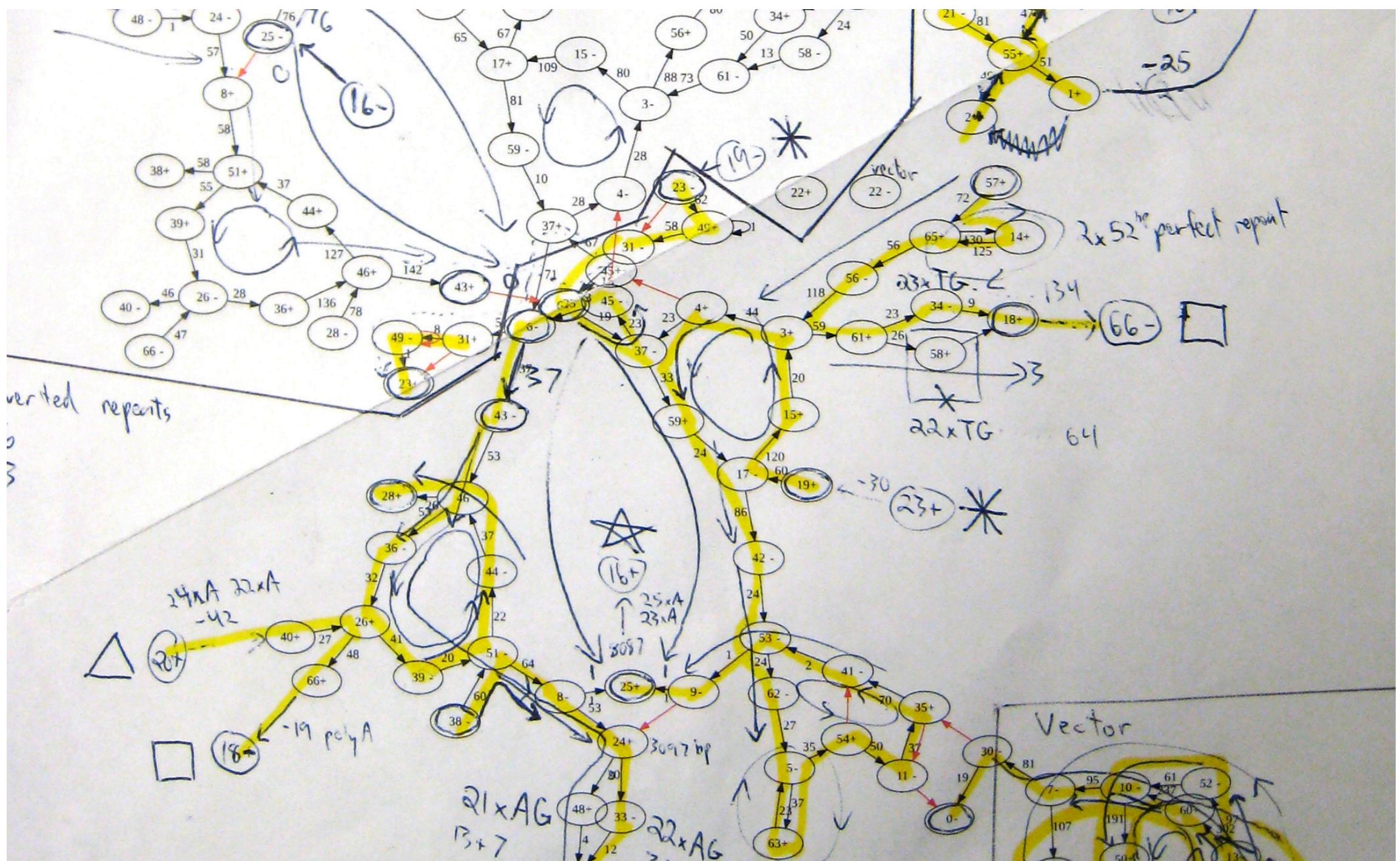


Image credit: Shaun Jackman, BC Cancer

What questions are you trying to answer?

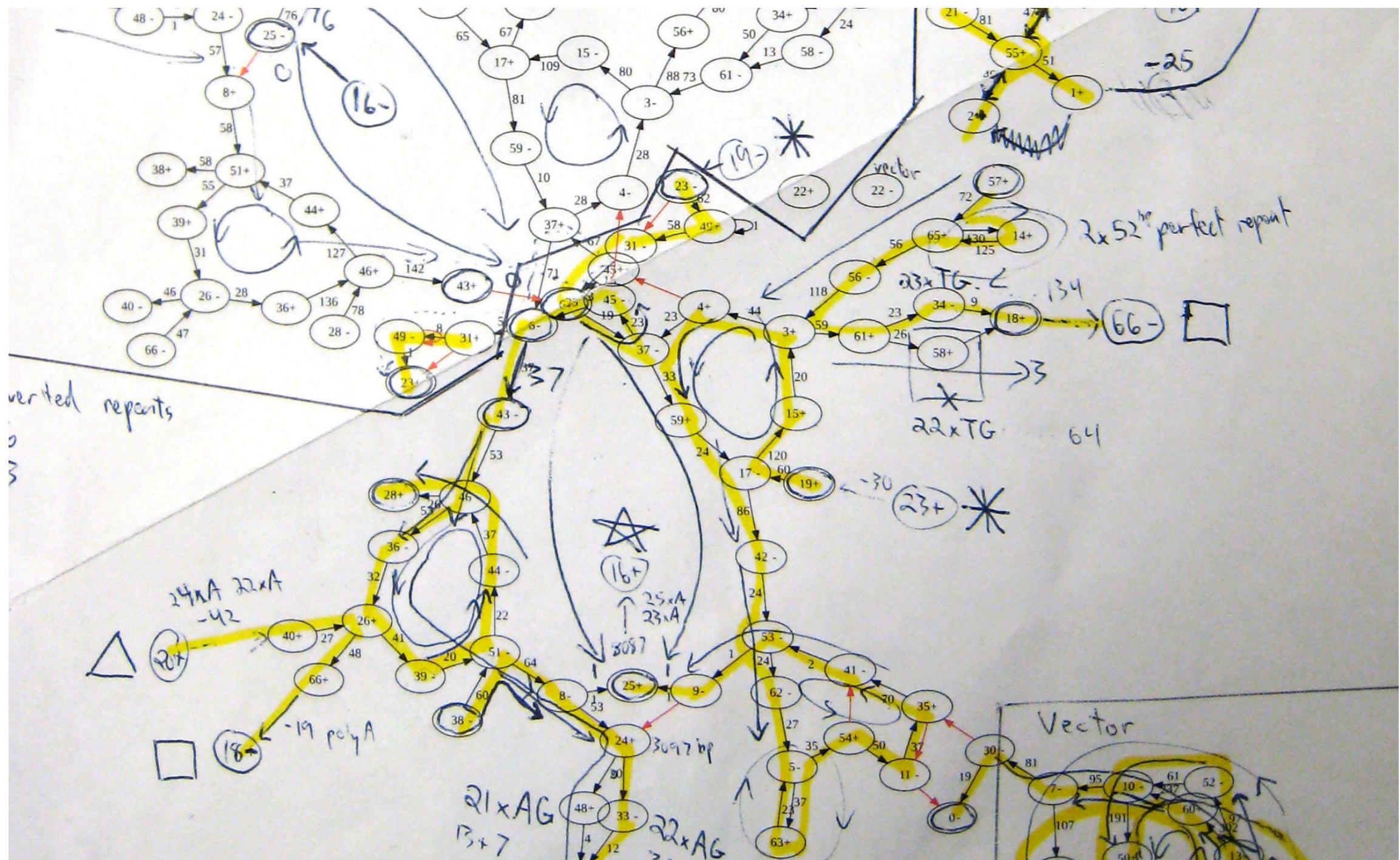


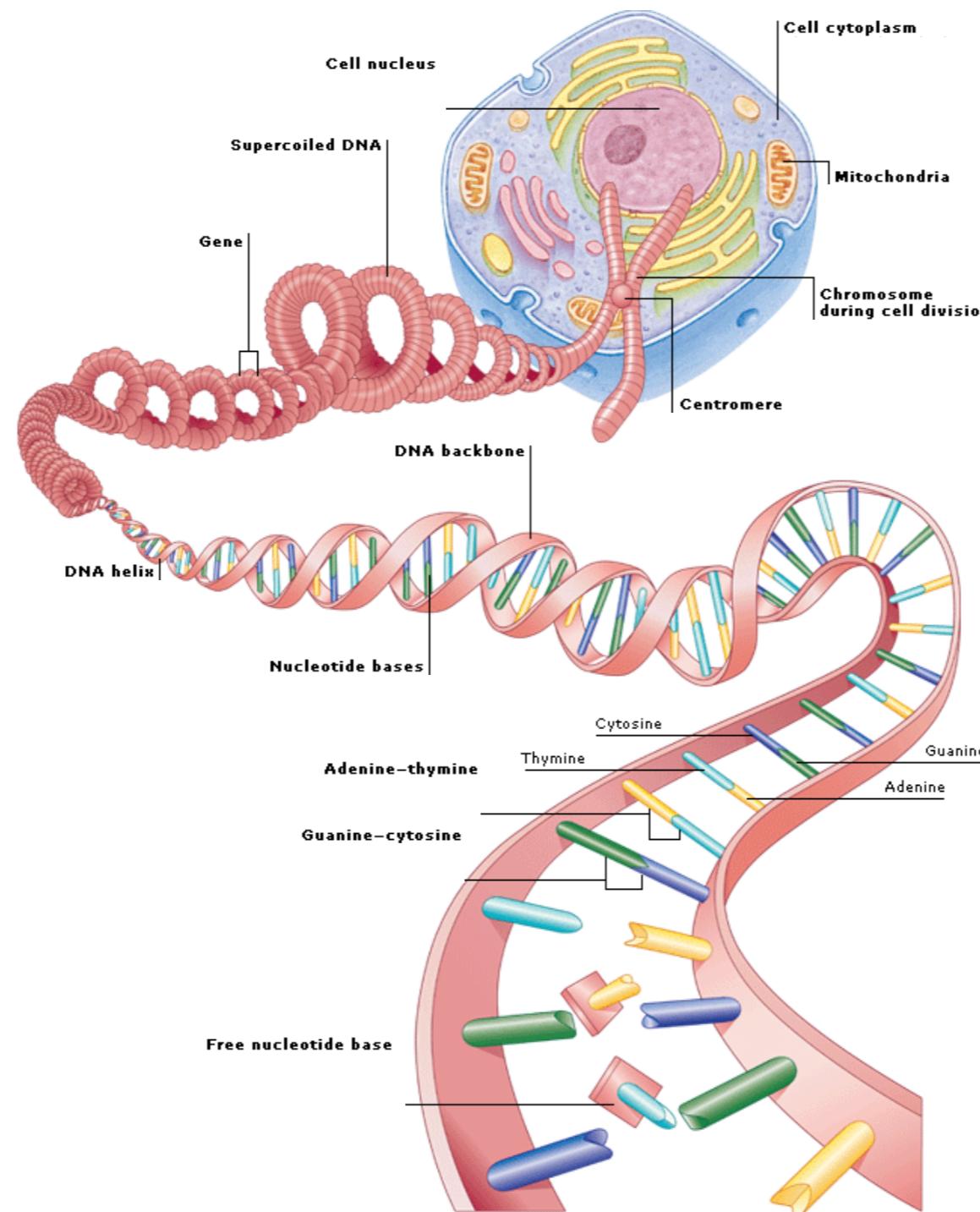
Image credit: Shaun Jackman, BC Cancer

Don't do data analysis in a void

Don't do data analysis in a void

- **Talk to the consumers of your analysis / visualization**
 - What do they care about?
 - What is their goal?
 - Do they want to make a decision? What do they need to make a decision?
 - Do they want evidence to support their claims?
 - Are they exploring the data?
- **Learn about the domain and analysis context**
 - The more you understand the problem, the more you will be able to think critically about it and contribute valuable insights
 - Let yourself be curious!

Real World Example | Genomics domain



Medical Encyclopedia
<https://www.aviva.co.uk>

We can read a genome in short sequences



Putting the short sequences back together

Assembly is challenging because of repetitive elements

True sequence

GGATTGAAAAAAGTACGCACGAATATACTAGCAAAAAATTACG

Sequencing data

GGATTGAAAAAA

TGAAAAAA

AAAAAAA

AAAAAAAGTAG

AAAGTCACGAATAT

AATATACATAGCAAA

AAAAAA

AAAAAAATTACG

Putting the short sequences back together

Assembly is challenging because of repetitive elements

True sequence

GGATTGAAAAAAGTACGCACGAATATACTAGCAGTTACG

Sequencing data

AAAAAA

AAAAAA

Putting the short sequences back together

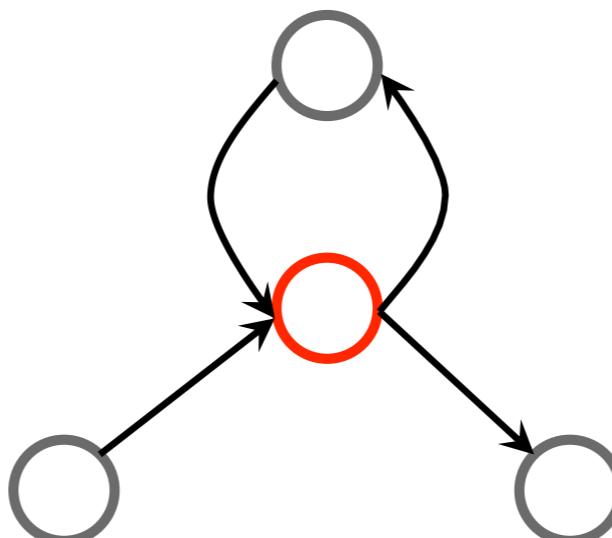
Assembly is challenging because of repetitive elements

True sequence

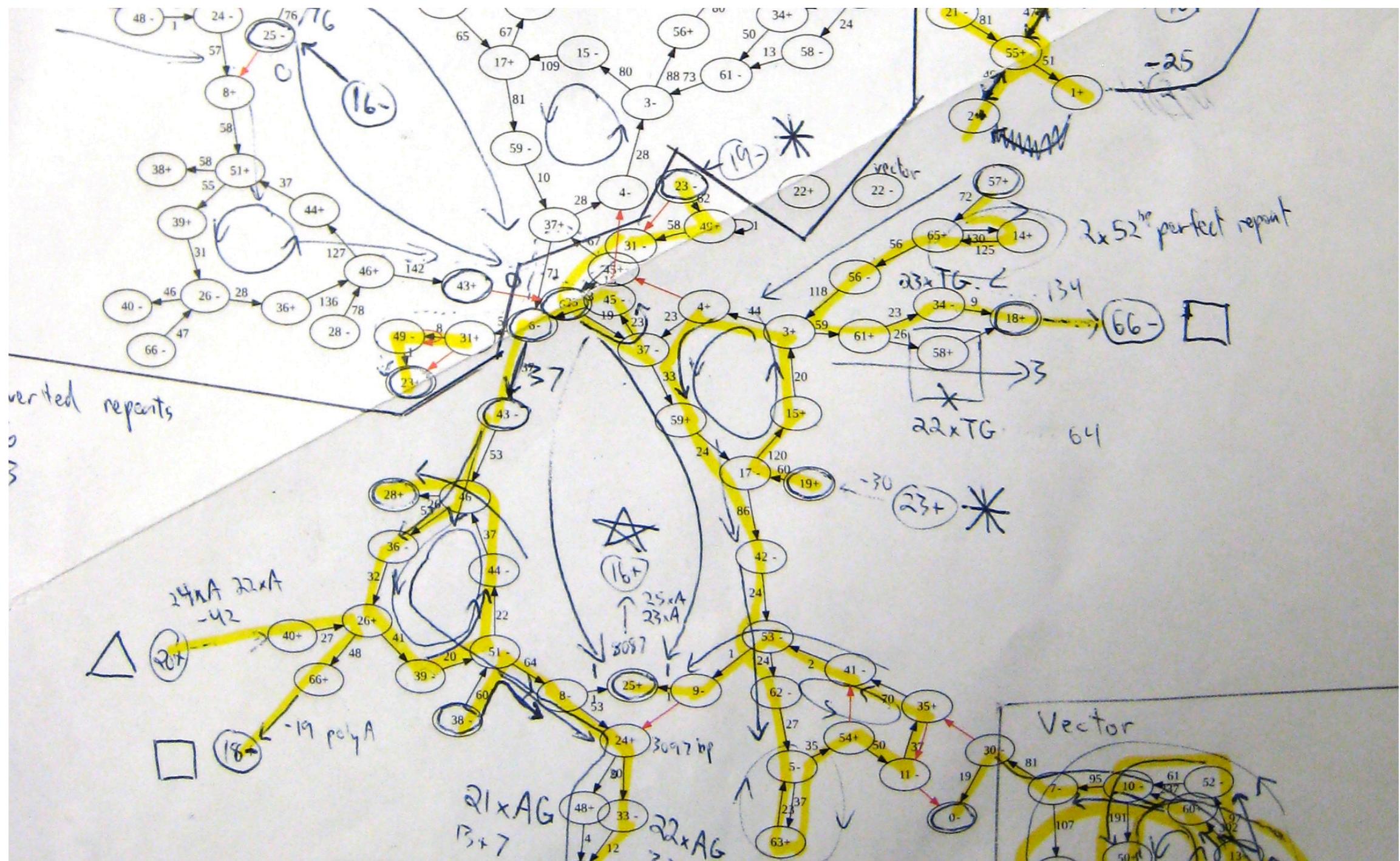
GGATTGAAAAAAGTACGCACGAATATACTACATAGCAAAAAATTACG

Assembled sequence graph

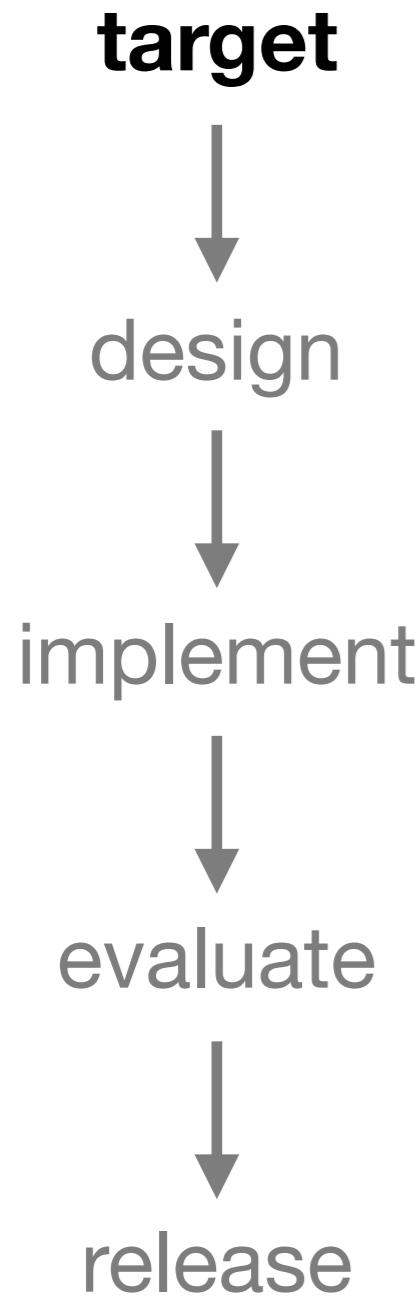
Assembly by Short Sequences (ABySS) algorithm



How are the sequences connected?



Target | Genome Vis Example



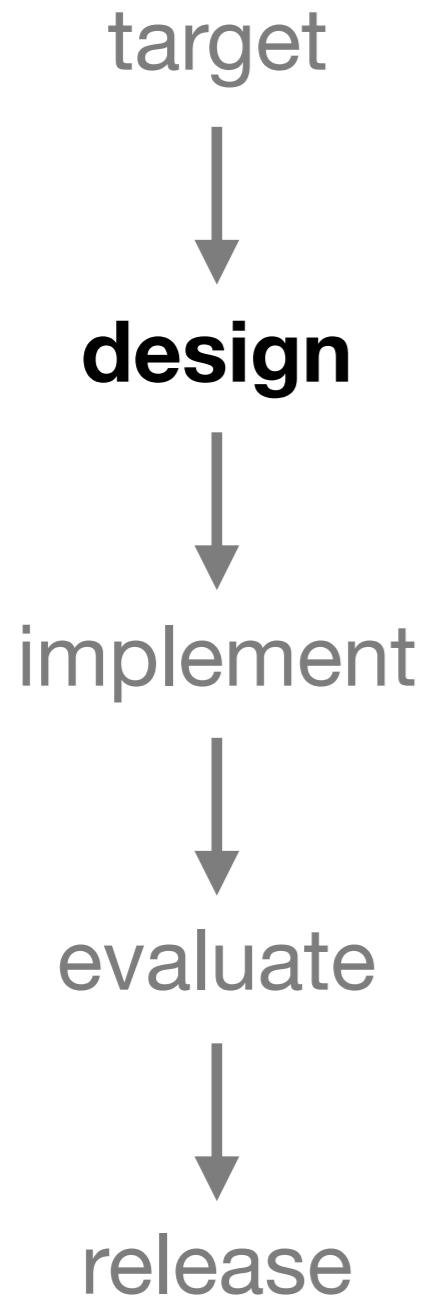
What is the purpose of my visualization?

- To help analysts assemble a genome correctly by:
 - Showing how the sequences are connected
 - Showing the lengths of the sequences (short versus long)
 - Showing whether the sequences are in the expected order
 - Revealing the neighbourhood of a query sequence

Who is my audience?

- Algorithm developers
- Data analysts

Design



What data should I show?

- Aim to show just what is needed for the target analytical task
- Consider the feasibility of showing large datasets

What data do I have?

- Sometimes when you start, the data you have isn't the data you need

How should I represent the data?

- Motivate your choice of visual encoding by considering the target analytical task

Should I make my visualization interactive?

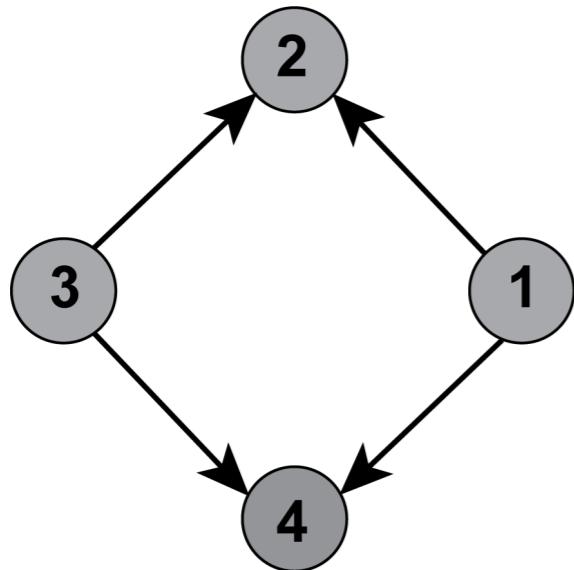
- As with the visual encoding, motivate your choice of interactivity by considering the target analytical task

Genome Vis Example | How are the sequences connected?

Genome Vis Example | How are the sequences connected?

Data I have

Raw data graph



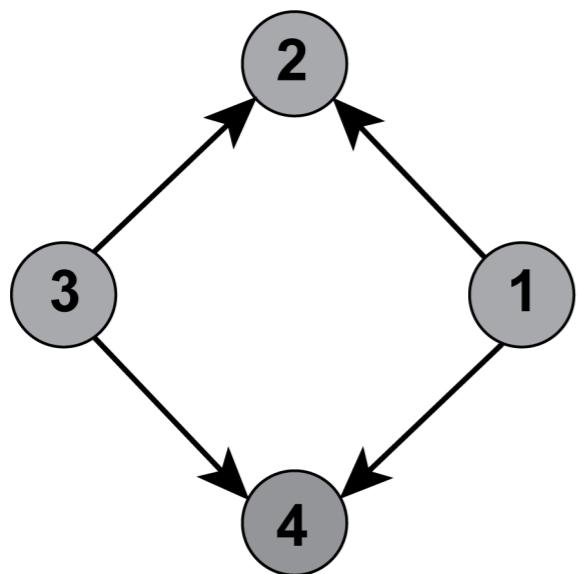
node = sequence

edge = $k-1$ overlap

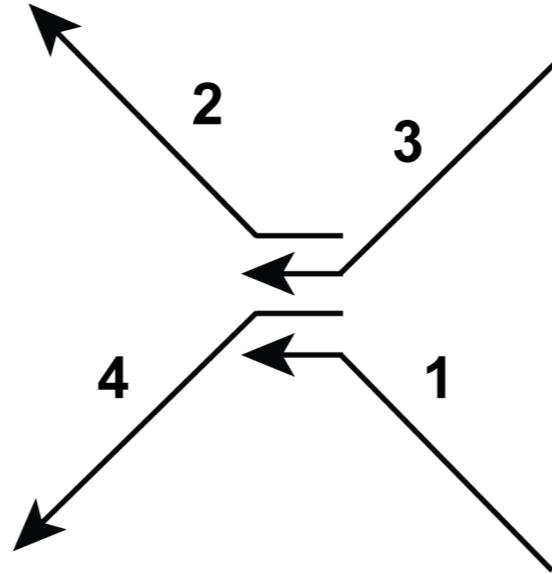
Genome Vis Example | How are the sequences connected?

Data I have

Raw data graph



DNA sequence

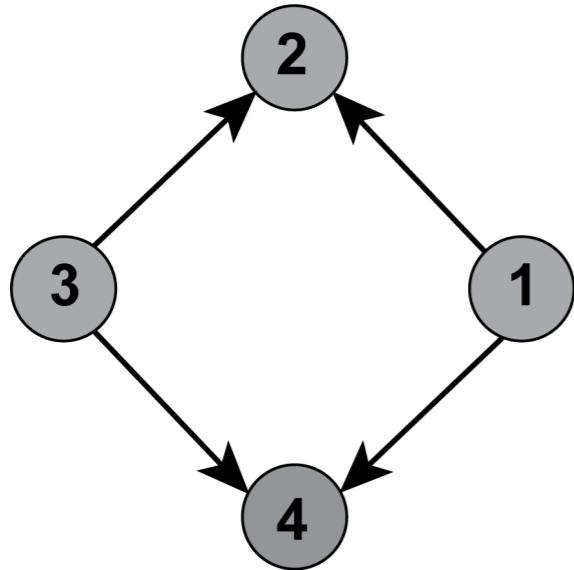


node = sequence
edge = $k-1$ overlap

Genome Vis Example | How are the sequences connected?

Data I have

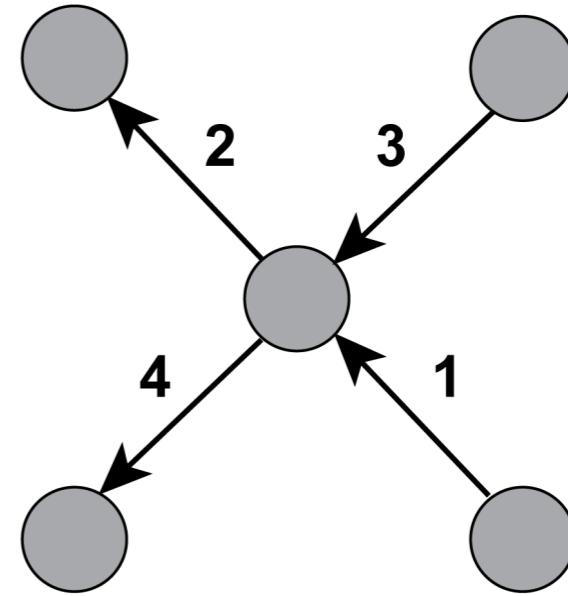
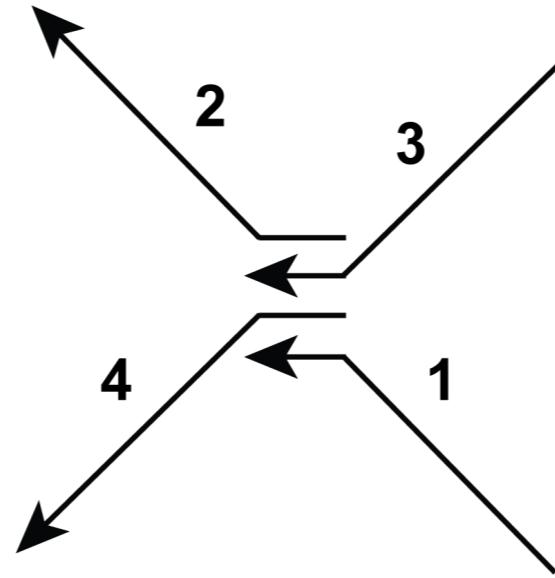
Raw data graph



node = sequence
edge = $k-1$ overlap

Data I want to show

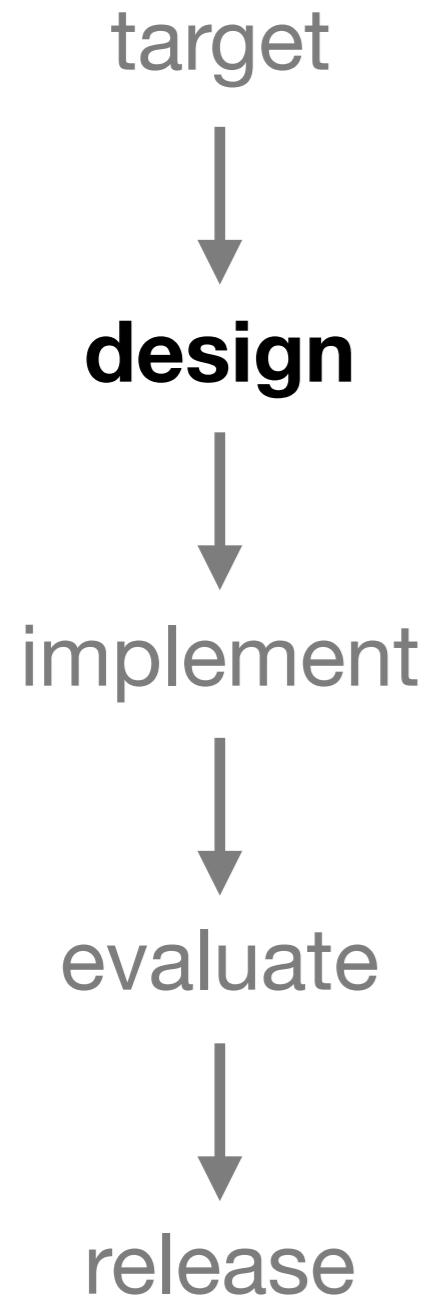
Transformed graph



node = $k-1$ overlap
edge = sequence

**Transform your data
to serve your analytical task**

Design



What data should I show?

- Aim to show just what is needed for the target analytical task
- Consider the feasibility of showing large datasets

What data do I have?

- Sometimes when you start, the data you have isn't the data you need

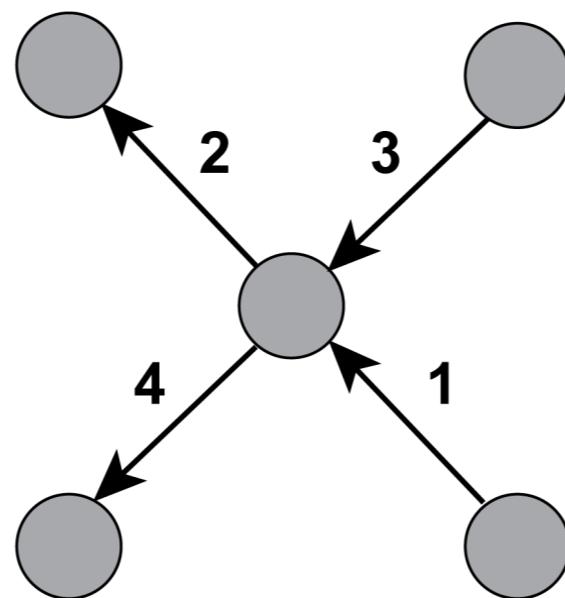
How should I represent the data?

- Motivate your choice of visual encoding by considering the target analytical task

Should I make my visualization interactive?

- As with the visual encoding, motivate your choice of interactivity by

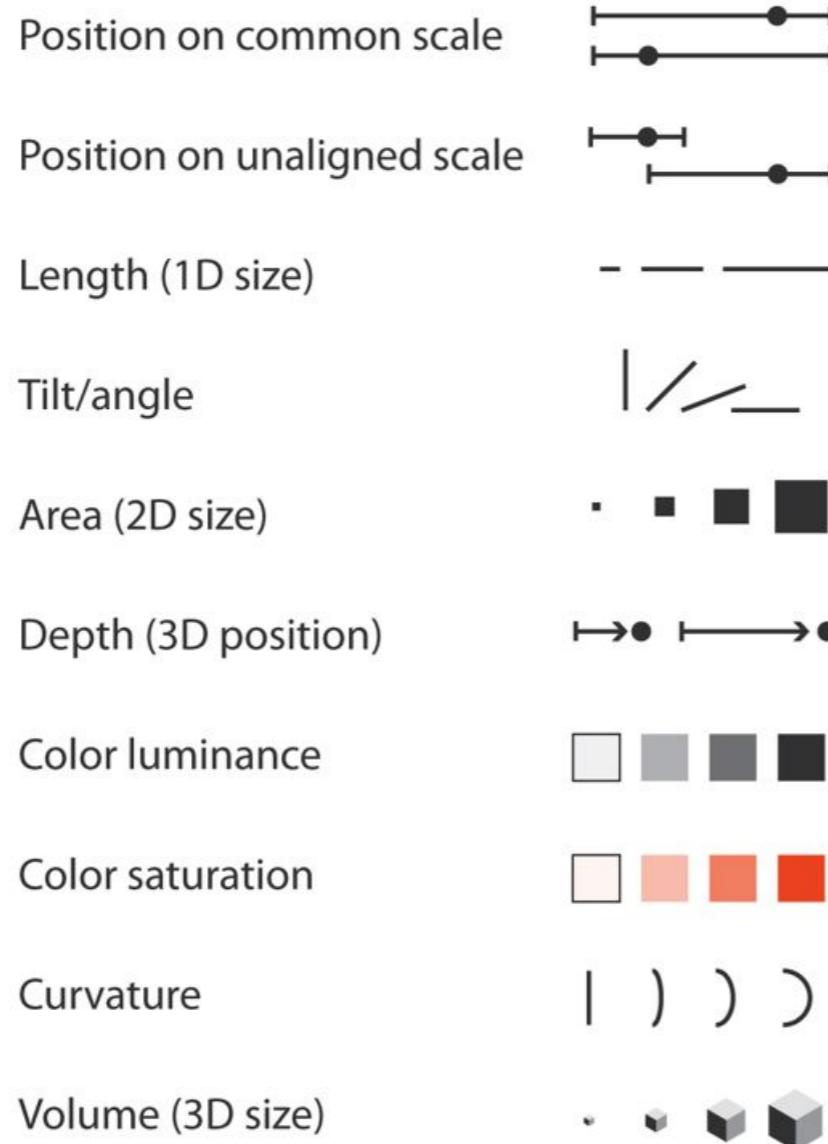
Genome Vis Example | How long are the sequences?



Principles | Choosing a visual encoding

Channels: Expressiveness Types and Effectiveness Ranks

④ **Magnitude Channels: Ordered Attributes**



④ **Identity Channels: Categorical Attributes**

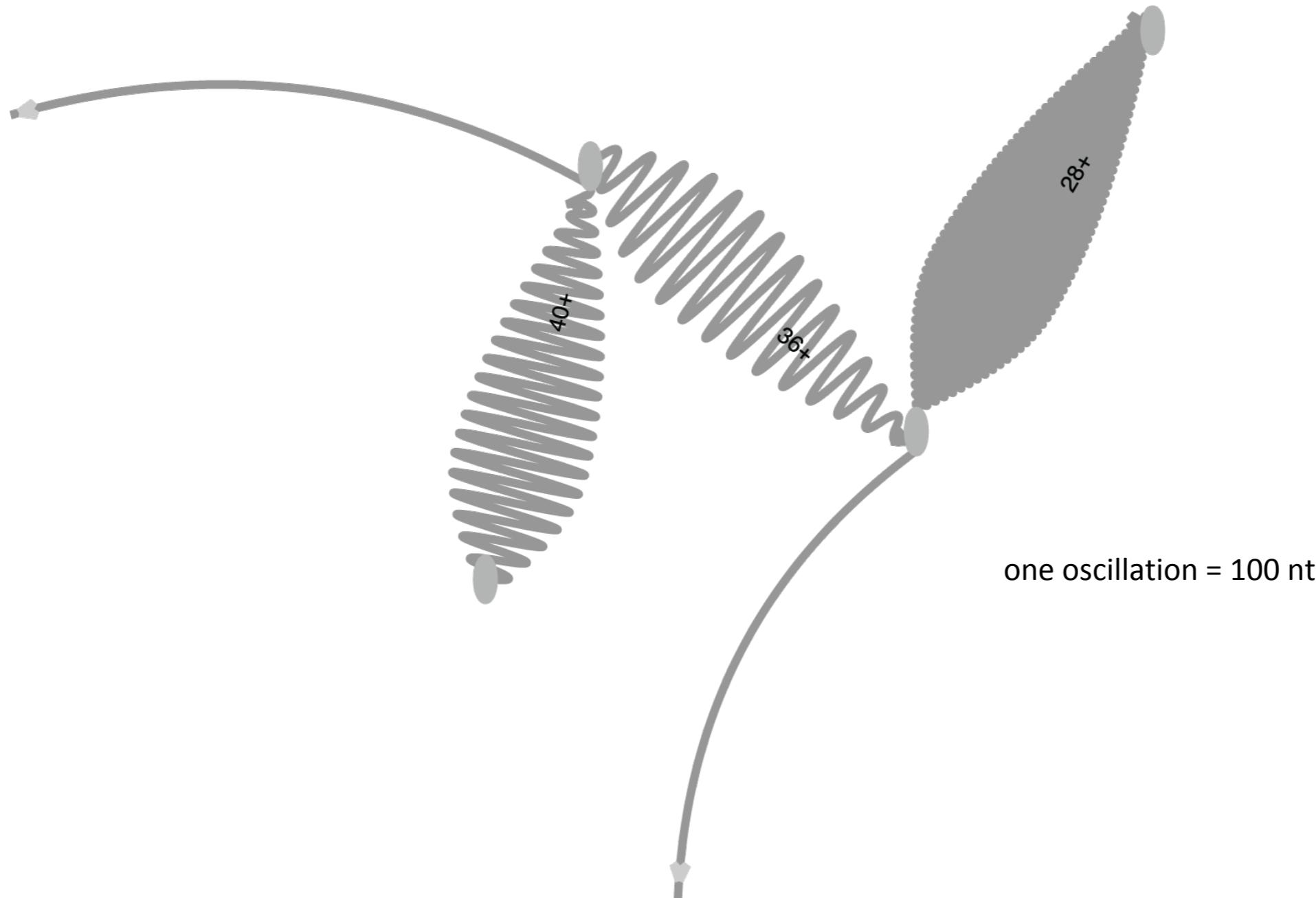


Get it right in black and white

- Maureen Stone

<http://www.stonesc.com/wordpress/2010/03/get-it-right-in-black-and-white/>

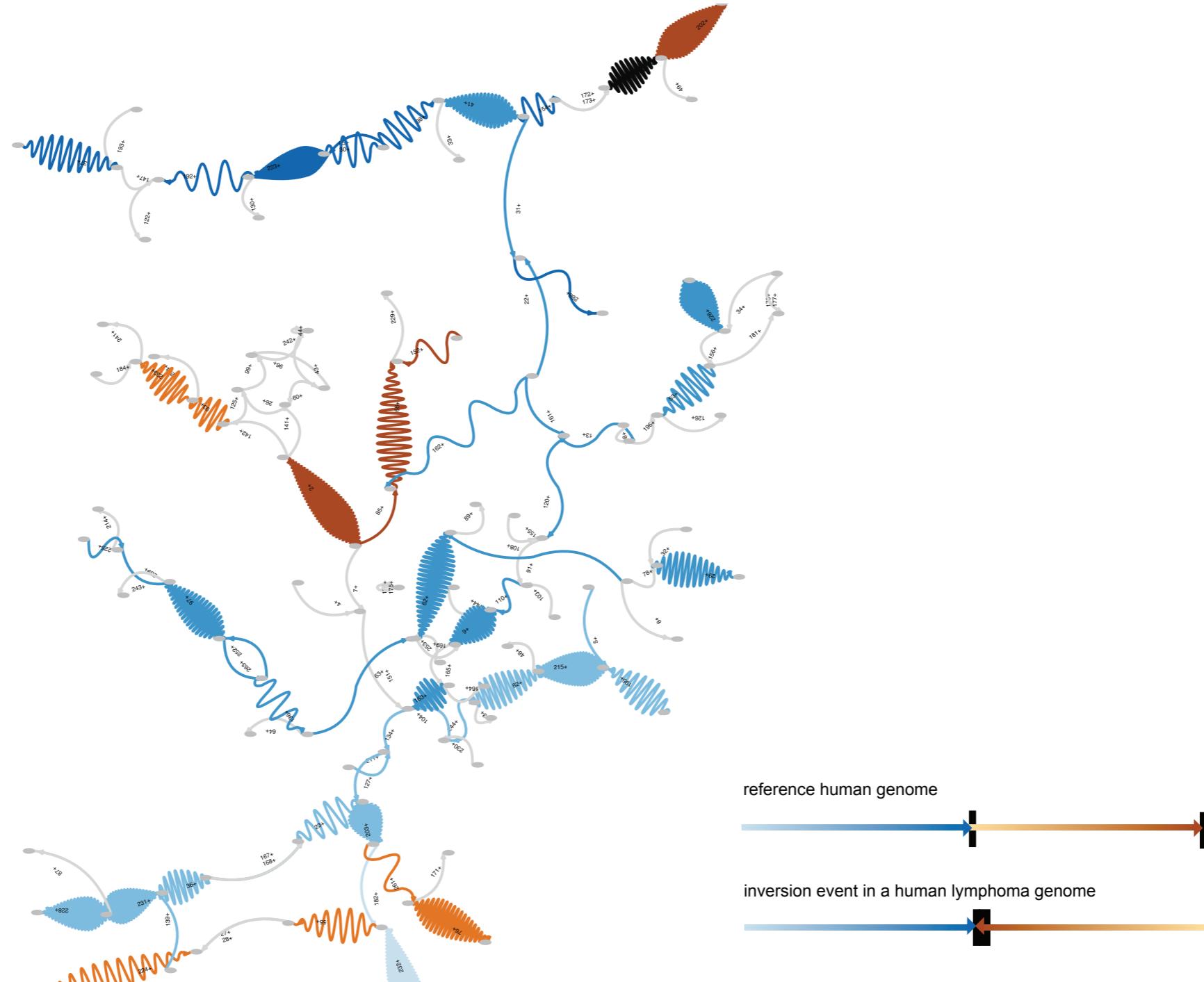
Genome Vis Example | How long are the sequences?



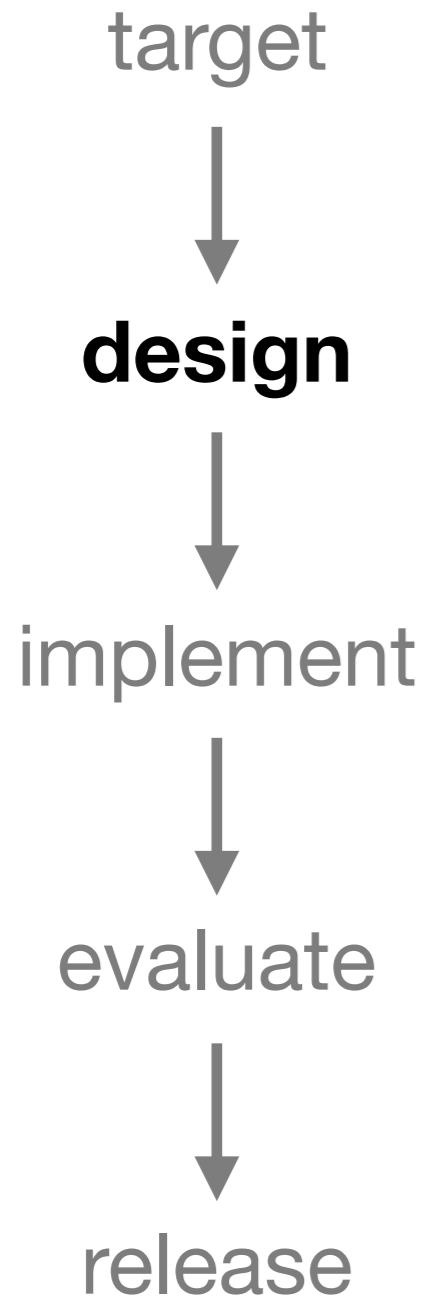
Genome Vis Example | How are they connected?



Genome Vis Example | Are they in the expected order?



Design



What data should I show?

- Aim to show just what is needed for the target analytical task
- Consider the feasibility of showing large datasets

What data do I have?

- Sometimes when you start, the data you have isn't the data you need

How should I represent the data?

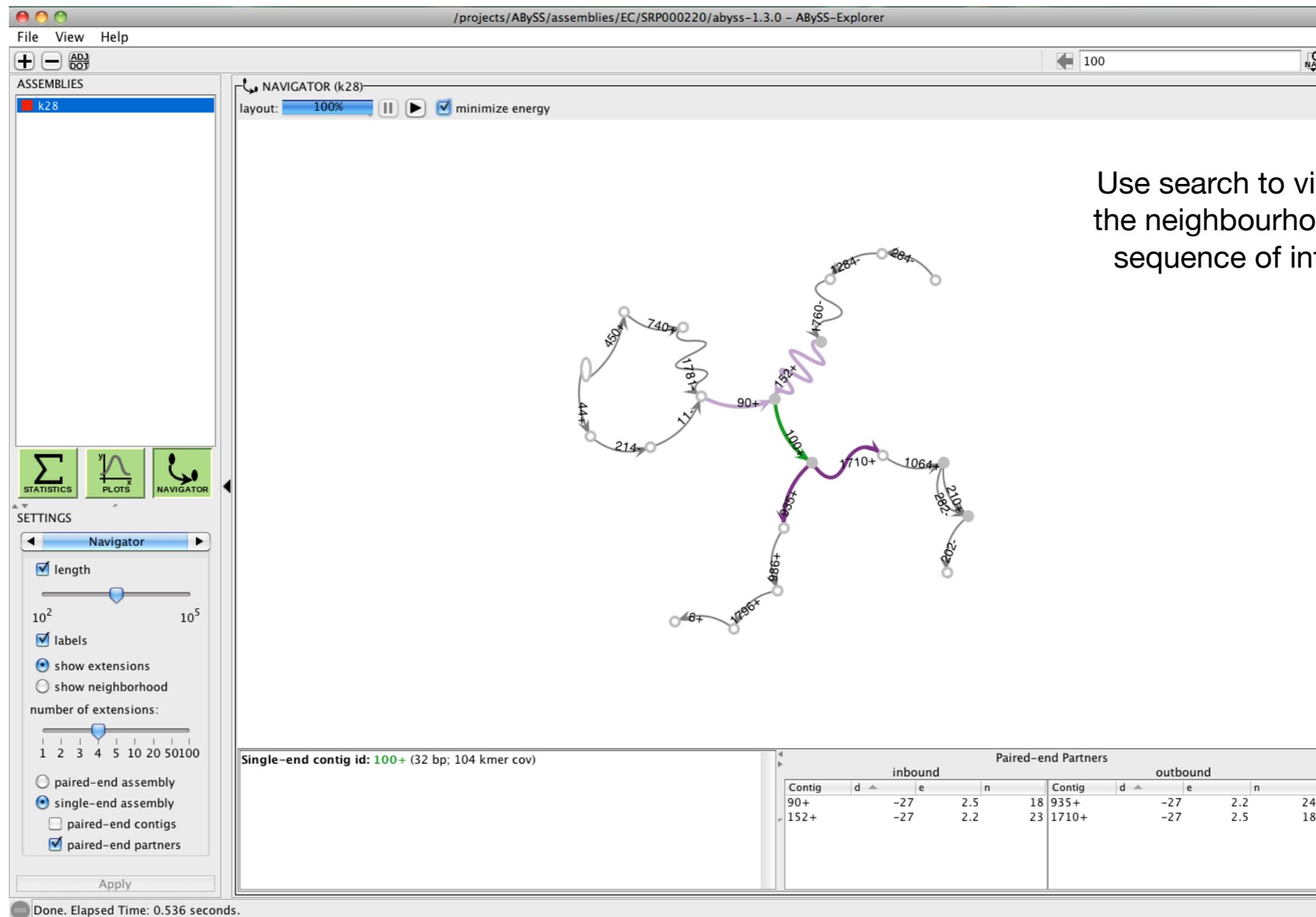
- Motivate your choice of visual encoding by considering the target analytical task

Should I make my visualization interactive?

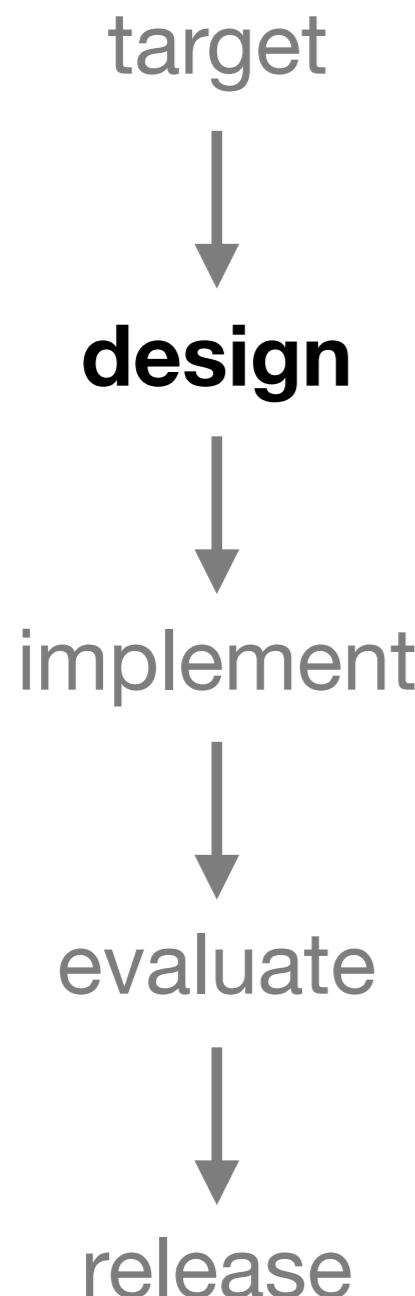
- As with the visual encoding, motivate your choice of interactivity by considering the target analytical task

**What people ask for may not
be what they want**

Genome Vis Example | Is this sequence correctly assembled?



Design | Genome Vis Example



What data should I show?

- Just show neighbourhood of a sequence of interest

What data do I have?

- Output of ABySS algorithm - first transform for visualization

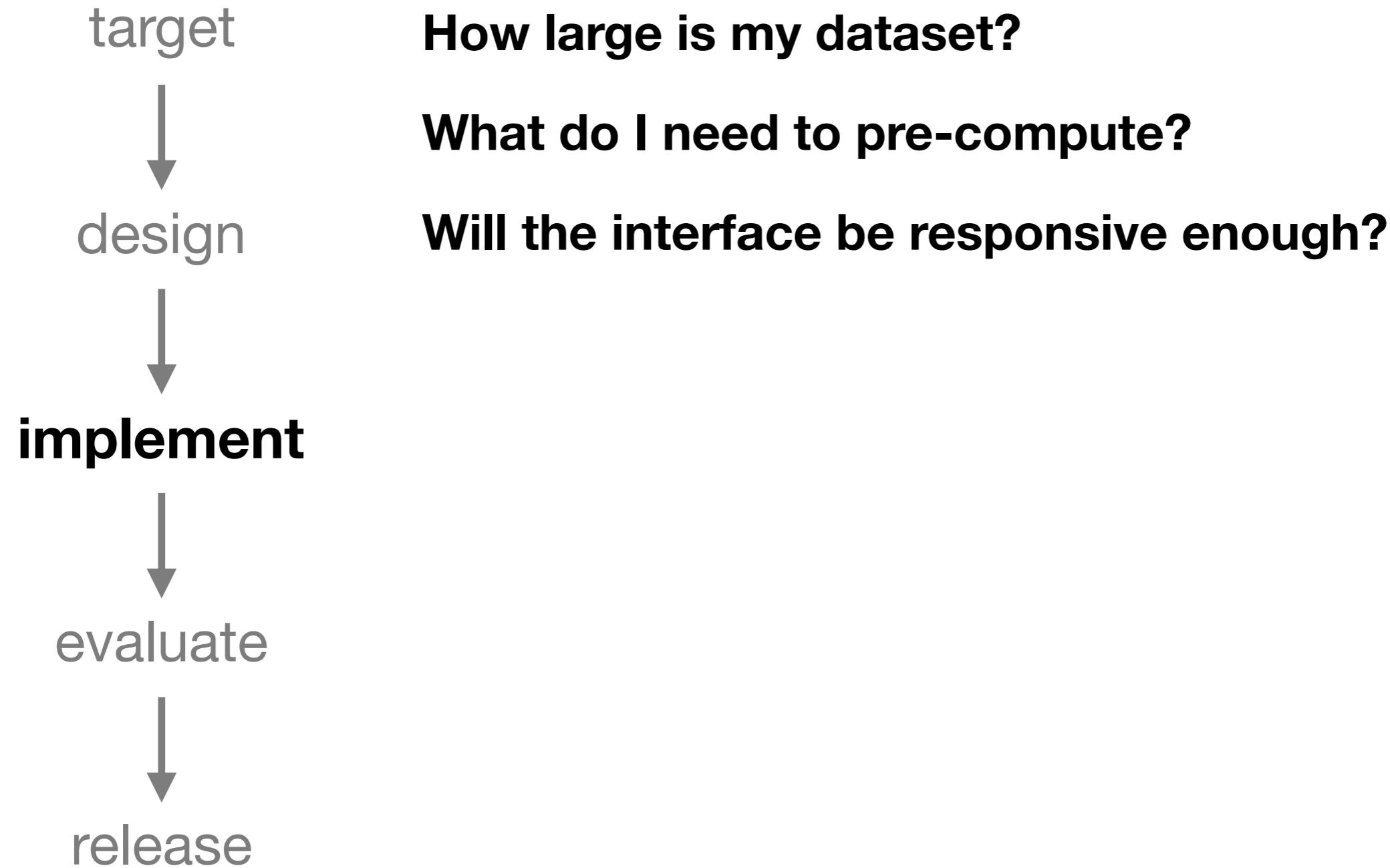
How should I represent the data?

- Represent as a graph (connectivity key)
- Show sequence lengths with waves
- Save colour for last (annotations on the core graph structure)

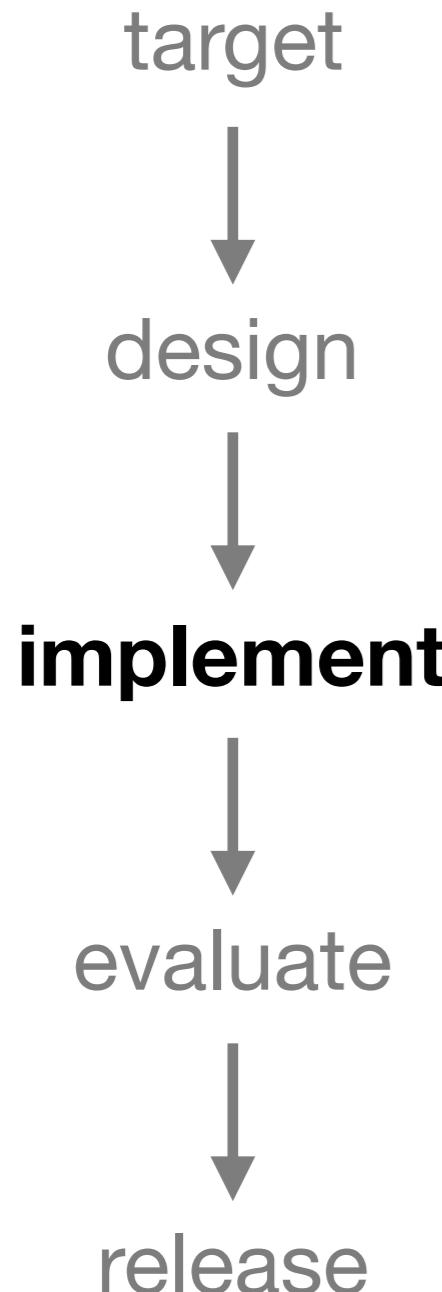
Should I make my visualization interactive?

- Dynamically draw the neighbourhood around a query sequence

Implement



Implement | Genome Vis Example



How large is my dataset?

- Hundreds to hundreds of thousands of sequences

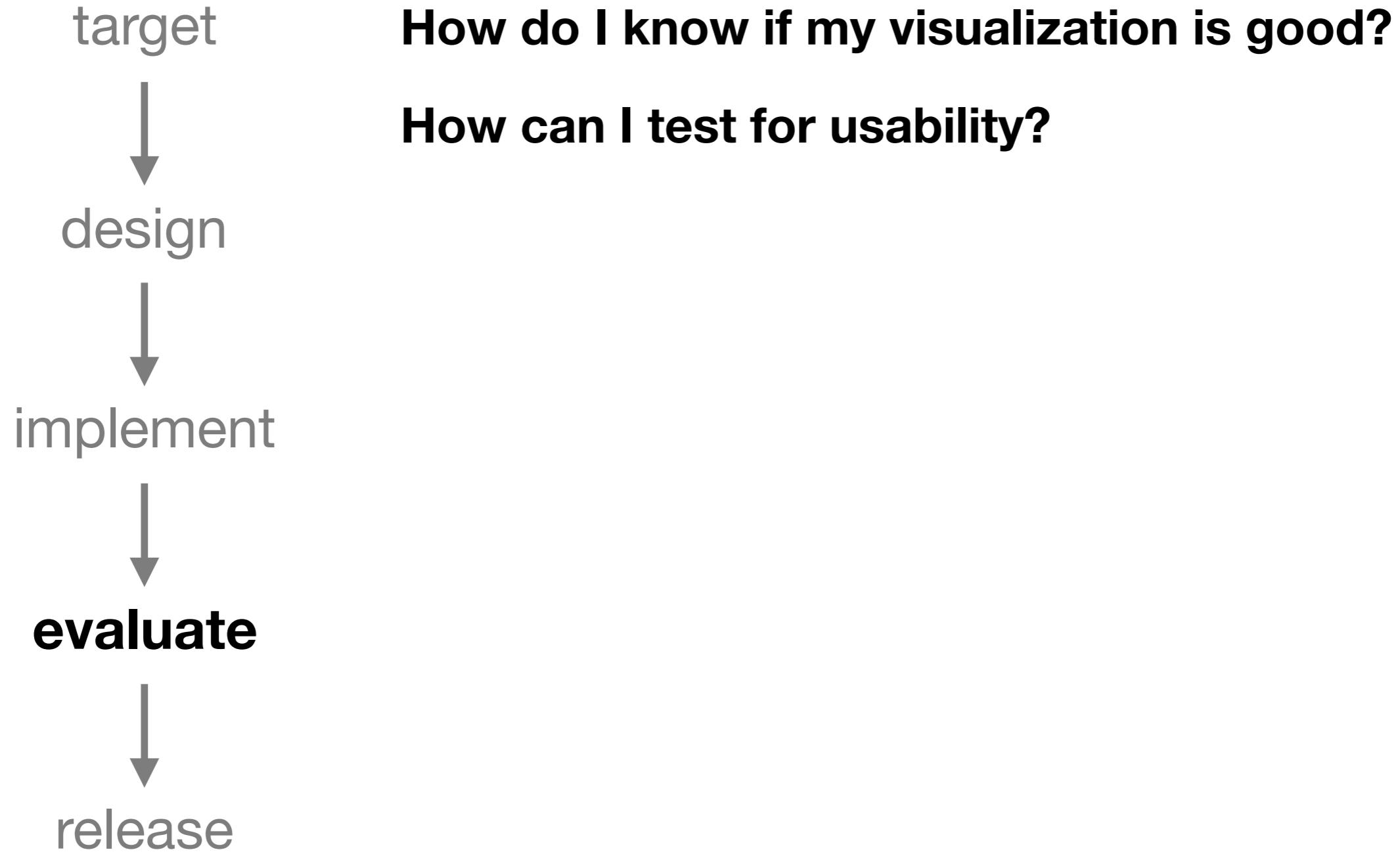
What do I need to pre-compute?

- Transform the output of ABySS to a graph for visualization

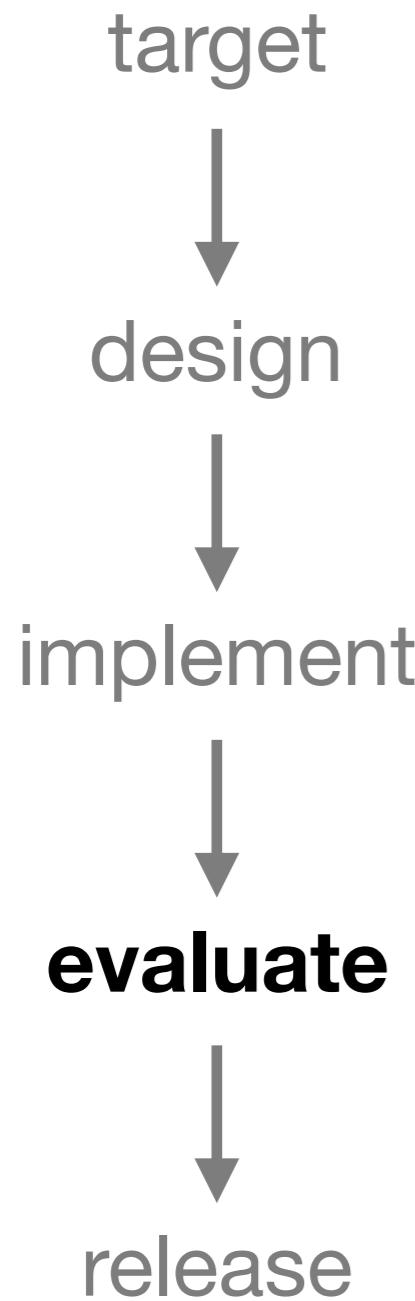
Will the interface be responsive enough?

- Displaying just a neighbourhood solved performance issues

Evaluate



Evaluate | Genome Vis Example



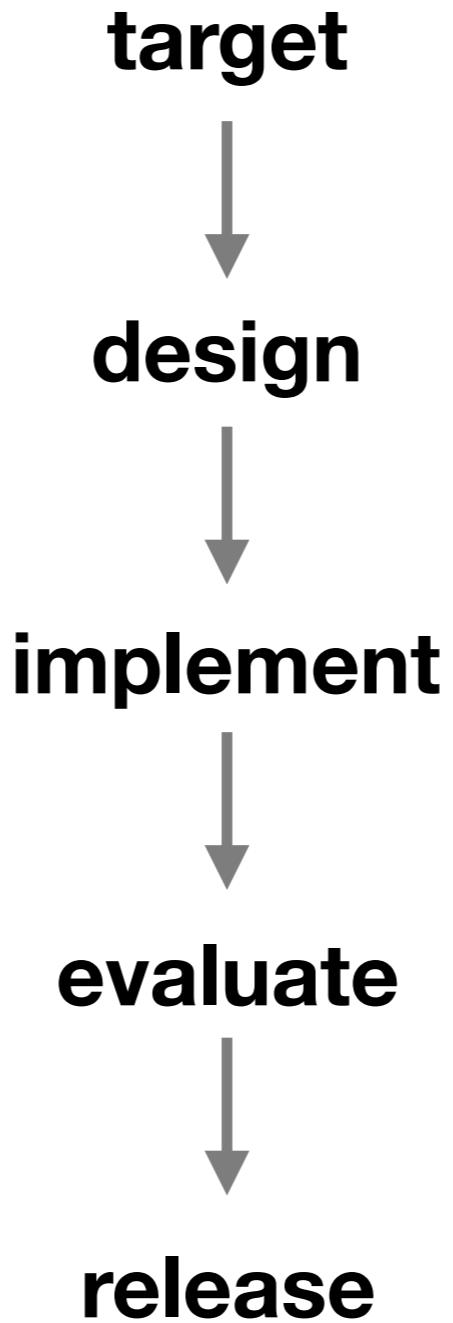
How do I know if my visualization is good?

- Interface is in active use for real world analysis

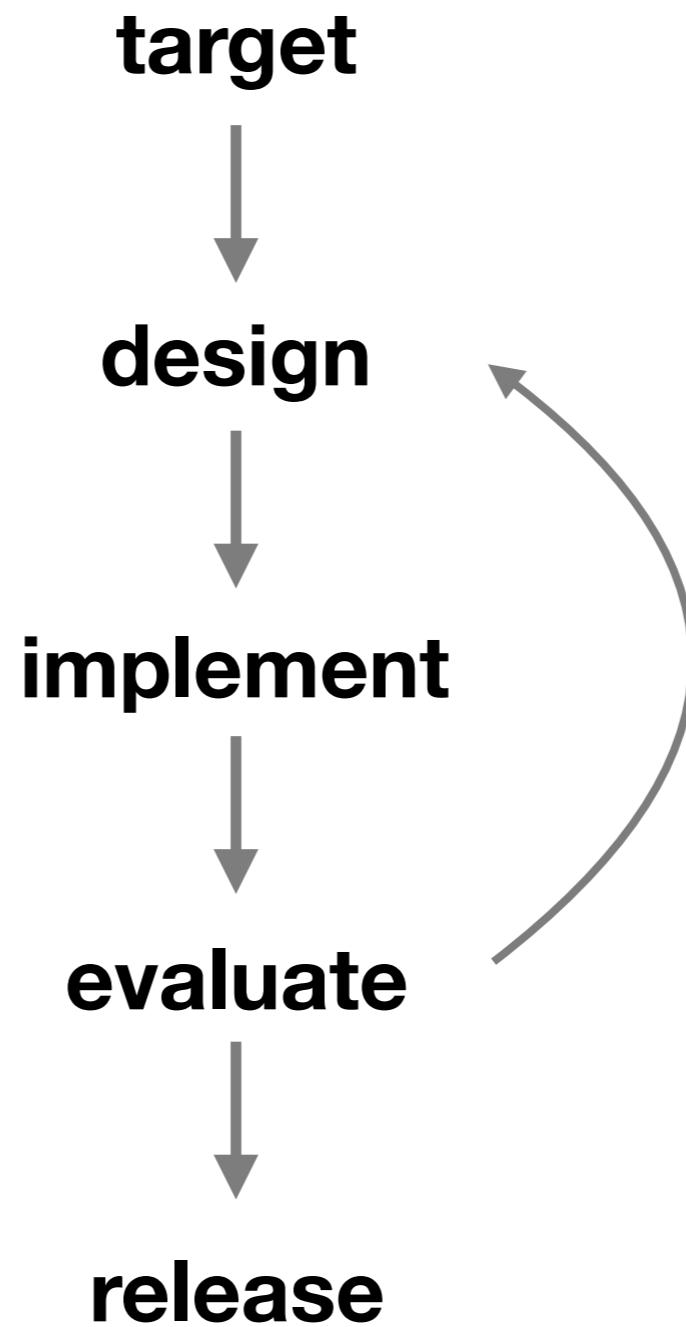
How can I test for usability?

- We worked closely with our users to iteratively test the functionality of the software throughout the design / development phases

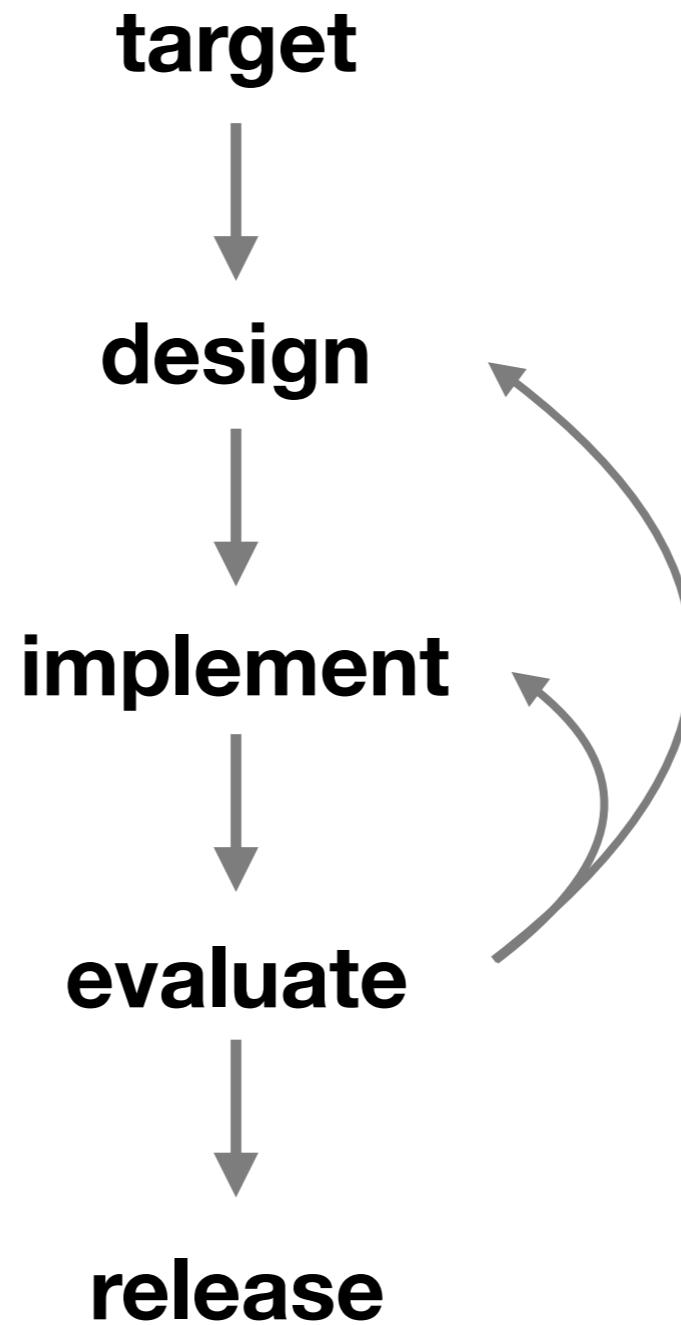
Iteration



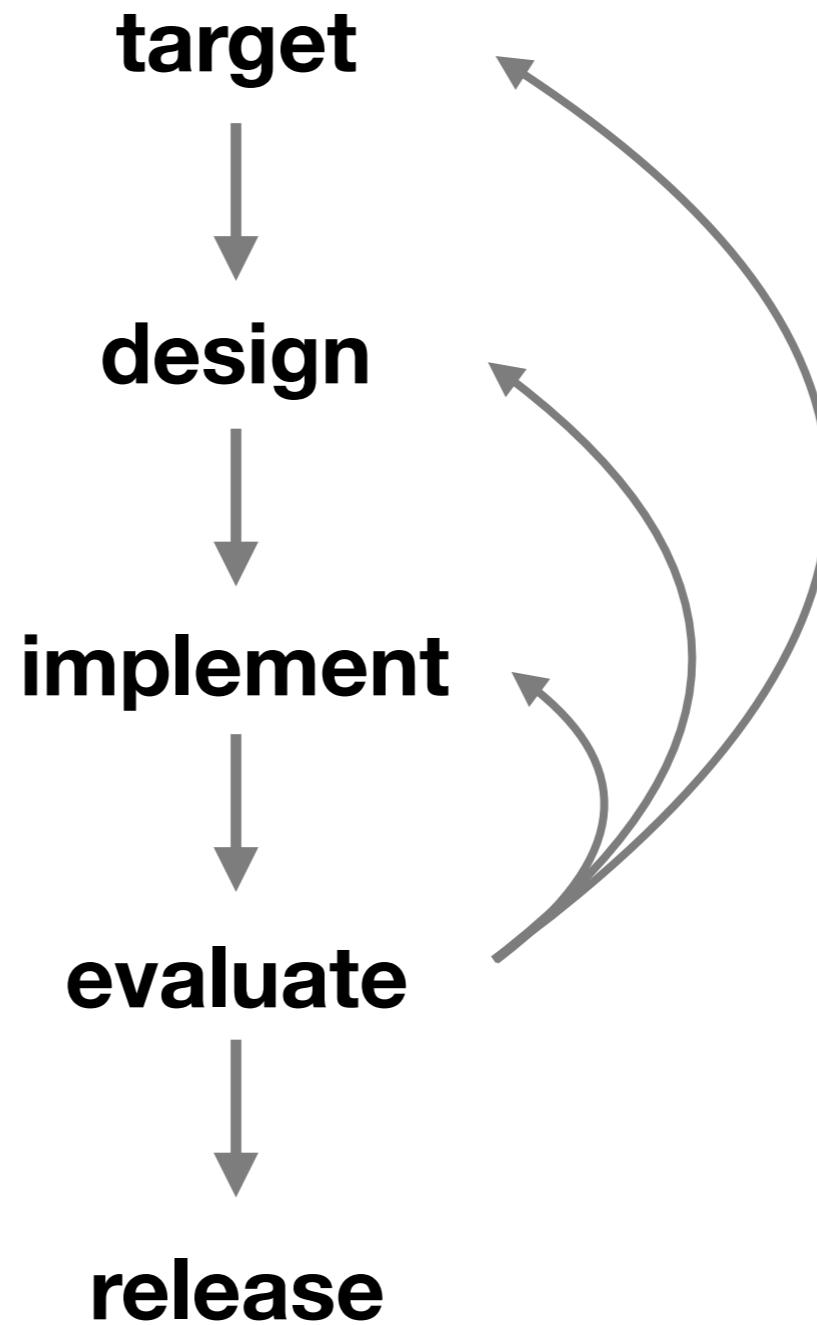
Iteration



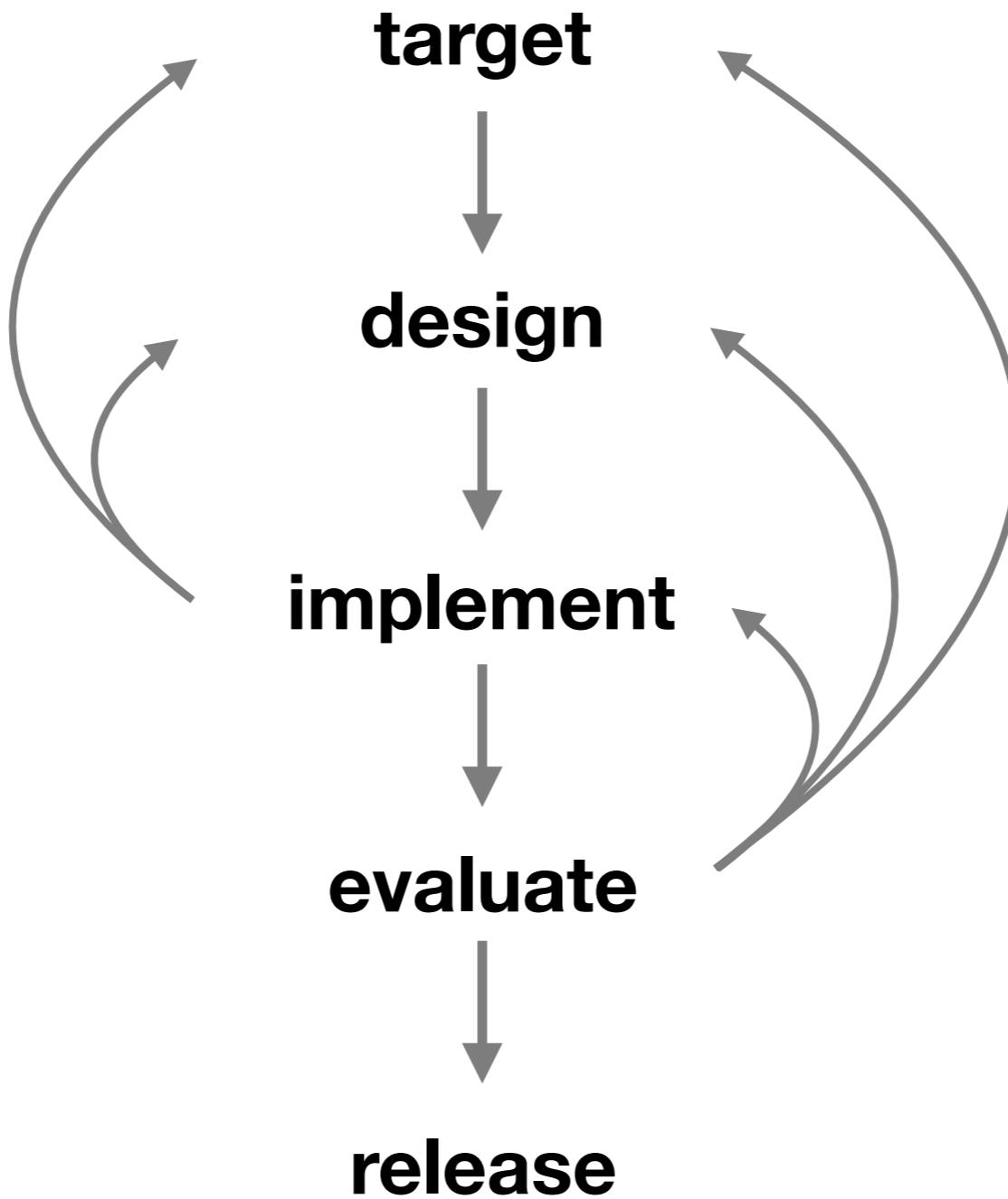
Iteration



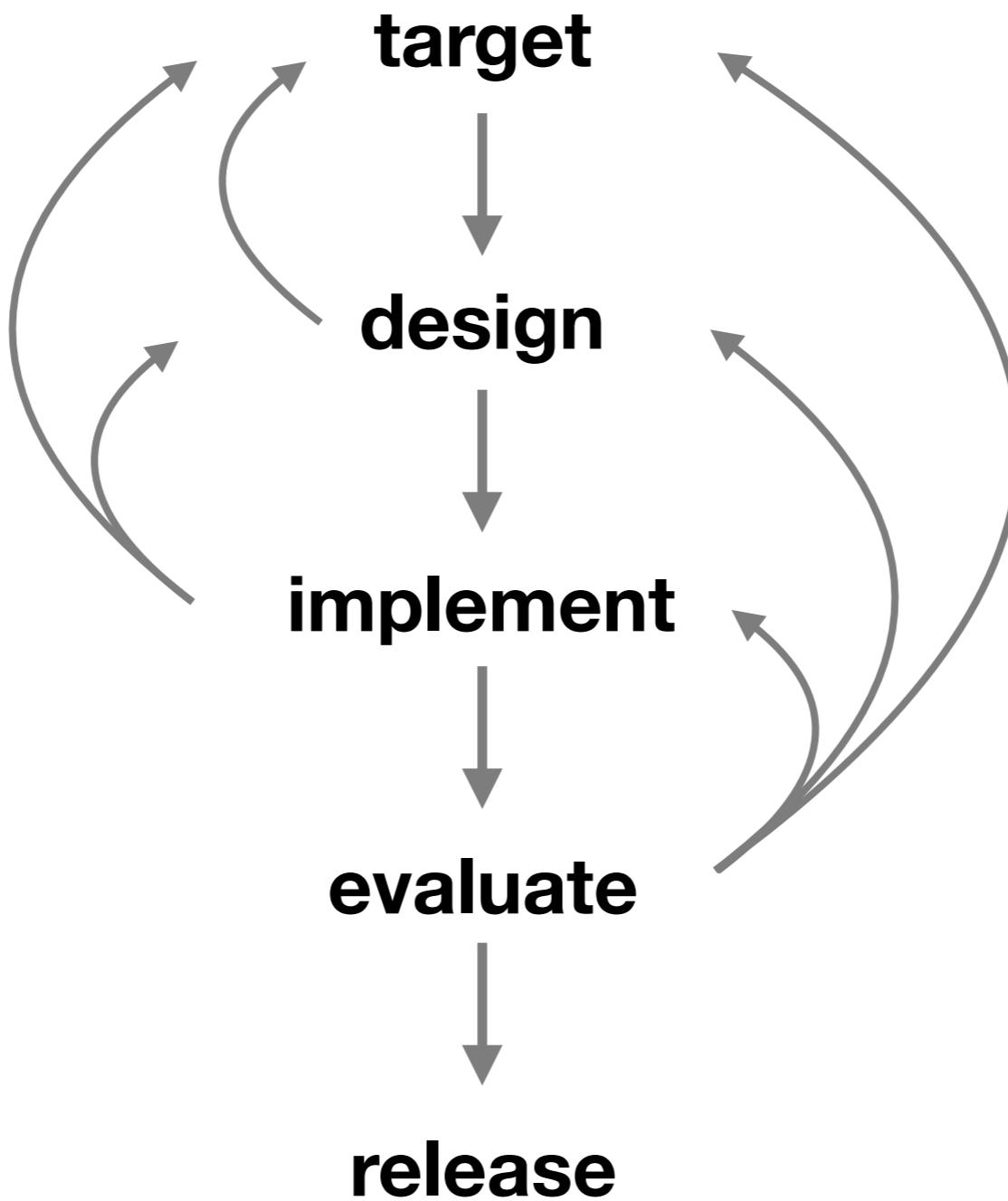
Iteration



Iteration



Iteration



Key points

- Start by clarifying the purpose of your visualization
- Transform your data to serve your analytical task
- Get it right in black and white
- What people ask for may not be what they want
- Think critically and stay curious!

Sneak peek

- | | |
|---|--------|
| 2. Introduction to Shiny Vincenzo Coia | Jan 7 |
| 3. Vis Principles and Best Practices | Jan 9 |
| 4. Interactivity - Manipulate | Jan 14 |
| 5. Interactivity - Facet | Jan 16 |
| 6. Interactivity - Reduce + Aggregate + Derive | Jan 21 |
| 7. Scalability | Jan 23 |
| 8. Usability + Review + Inspiration | Jan 28 |

References

- **ABySS-Explorer: Visualizing Genome Sequence Assemblies**
 - Paper | [http://cydney.org/Publications files/Nielsen_infovis_2009.pdf](http://cydney.org/Publications_files/Nielsen_infovis_2009.pdf)
 - Project code | <https://github.com/bcgsc/ABySS-explorer>
- **Design Study Methodology: Reflections from the Trenches and the Stacks**
 - Paper | <https://www.cs.ubc.ca/labs/imager/tr/2012/dsm/#paper>
- **Get it right in black and white**
 - Blog | <http://www.stonesc.com/wordpress/2010/03/get-it-right-in-black-and-white/>
- **Visualization Analysis & Design | Tamara Munzner**
 - Chapter 5 - Marks and Channels