

Data Visualization for Biomedical Applications

Lecture 6

BMI706 - 26 April 2018

Nils Gehlenborg, PhD

Administrative

- Please send us your presentation slides: bmi706.2018@gmail.com
- Submission deadline for **second project assignment**: Thursday, April 26 at 11:59 pm Eastern
- Additional Office Hours: Friday, April 27, from 2:00 to 3:00 pm in Countway 308

Administrative

- **Due April 19**

Data set if you are allowed to share it.

Description of data set with tasks.

- **Due Today**

Five design sheets.

Description of implementation strategy.

- **Due May 3**

Complete process book: merge everything into a *single document!*

Source code for Shiny app.

Presentation with live demo.

Final Presentation

- Timing: **10 minutes total = 7 minutes talk + 3 minutes for questions**
- Describe your data set and provide rationale for a visual exploration tool
- You should also briefly explain what computational techniques you used to support this process (clustering, etc.)
- Demo
- Discuss your final design and justify your design decisions
- Feedback and questions

Review of Session 4

Graph Visualizations

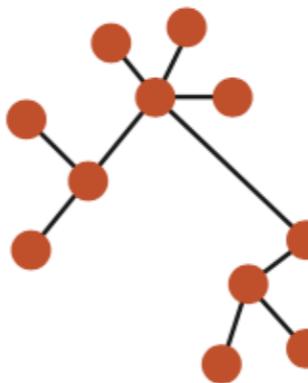
Arrange Networks and Trees

→ Node–Link Diagrams

Connection Marks

NETWORKS

TREES



→ Adjacency Matrix

Derived Table

NETWORKS

TREES



→ Enclosure

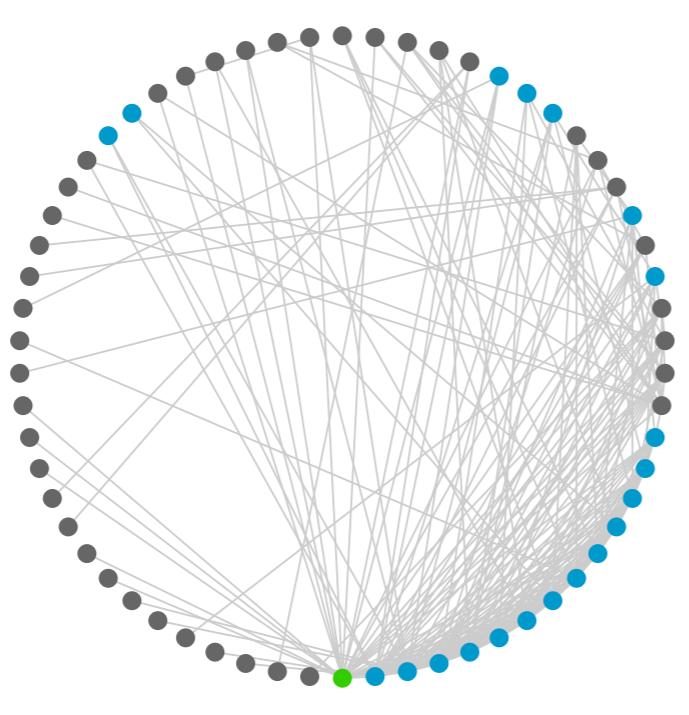
Containment Marks

NETWORKS

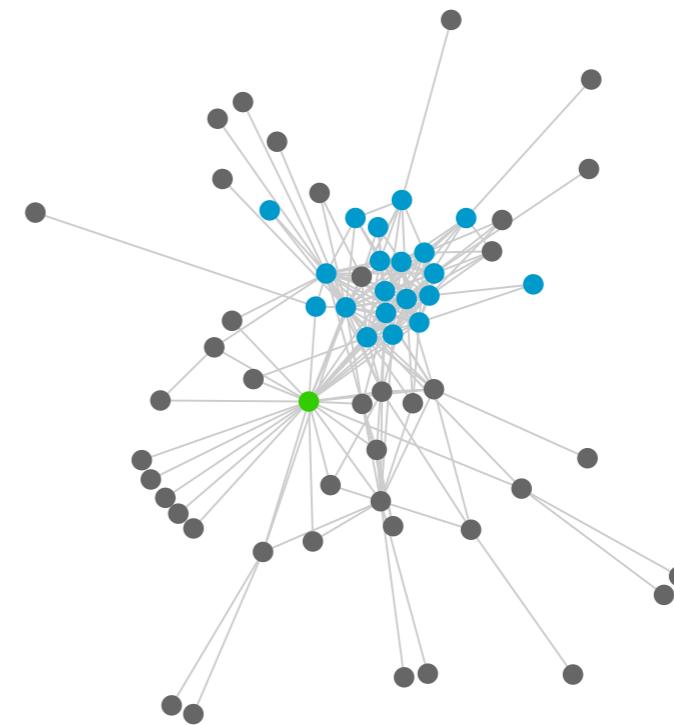
TREES



Layouts

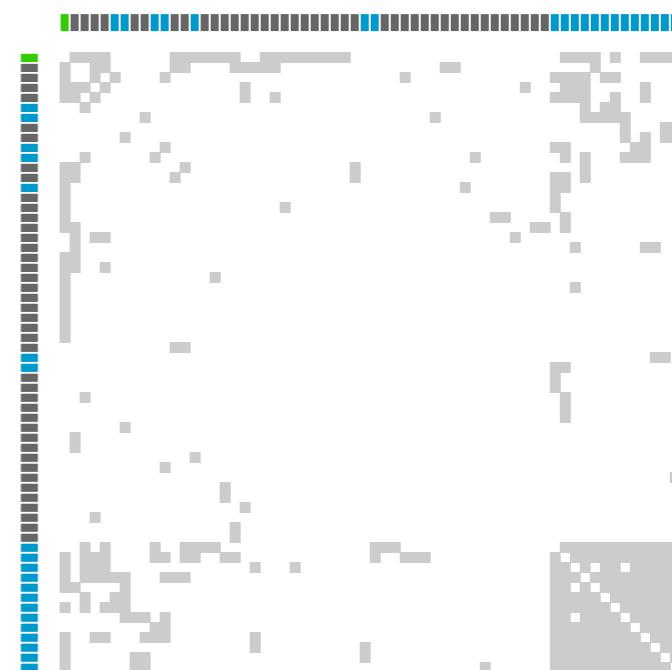
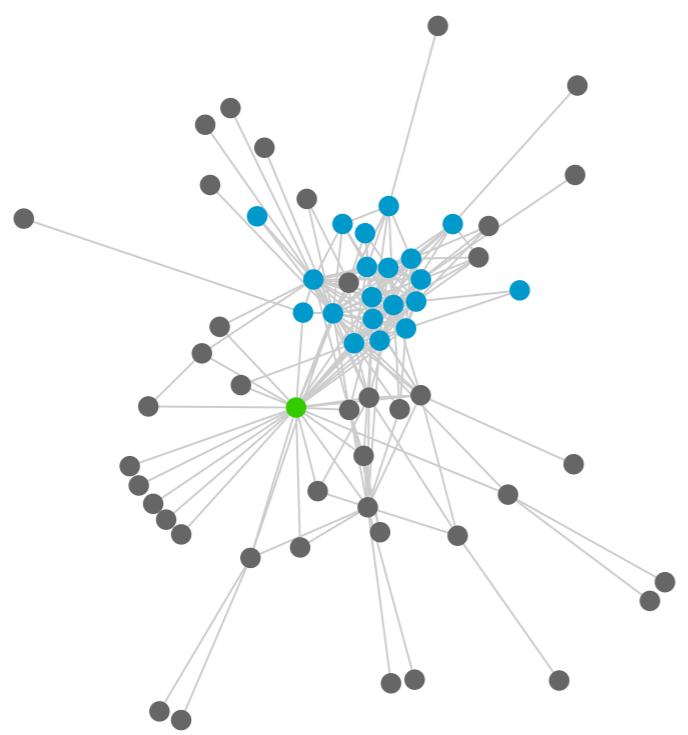


Circular

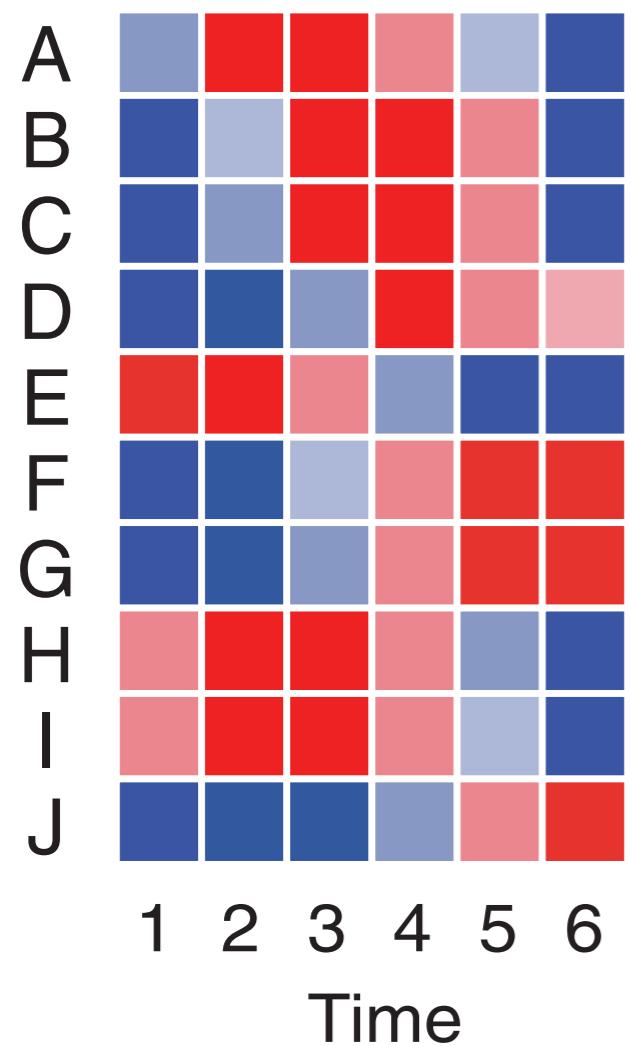


Force-Directed

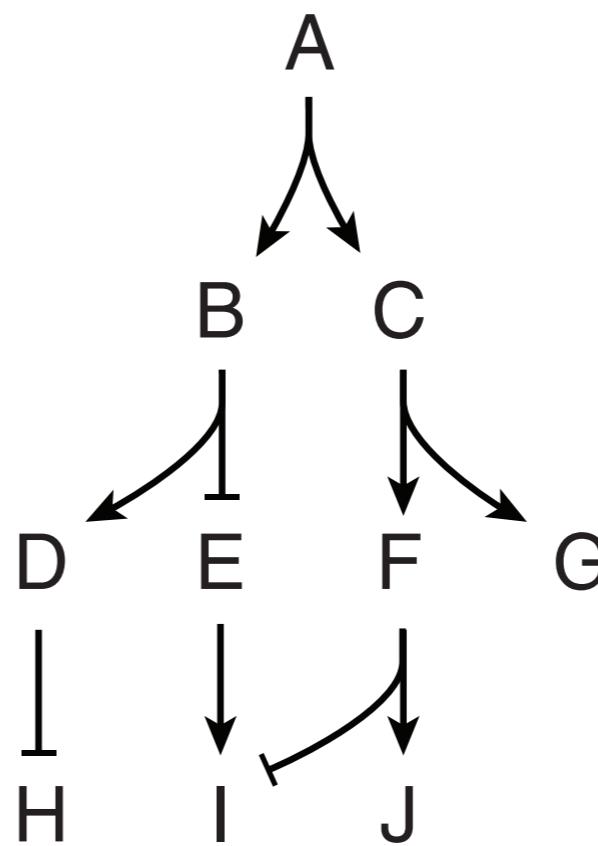
Pros and Cons



Multivariate Data Integration

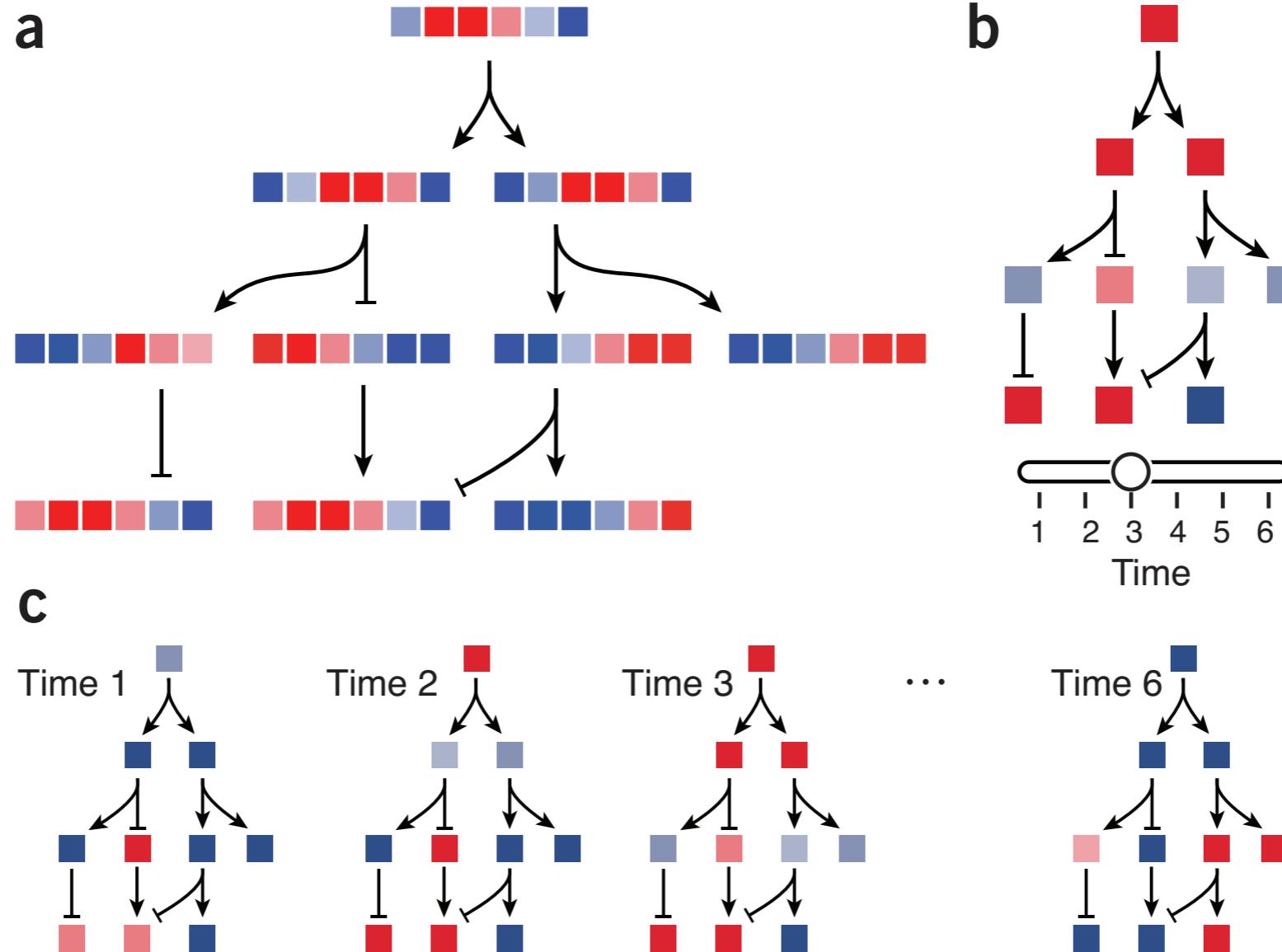


+

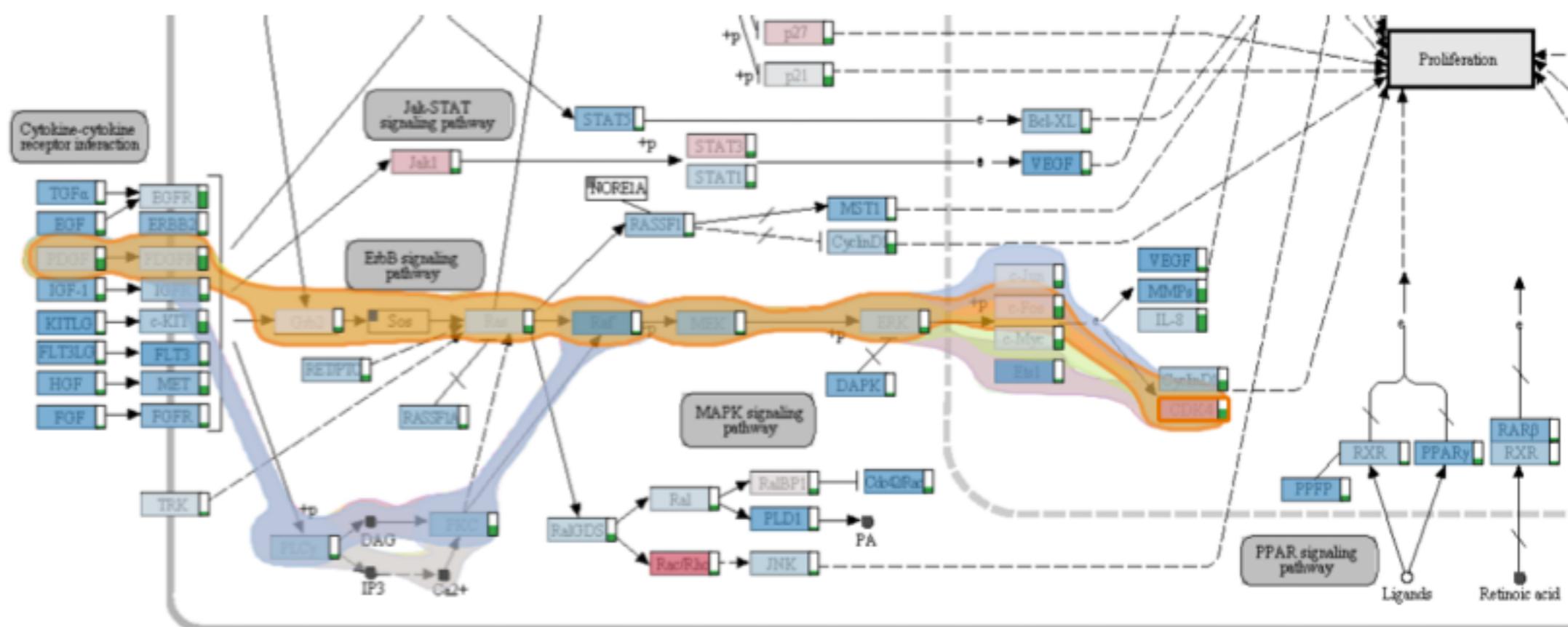


=

?



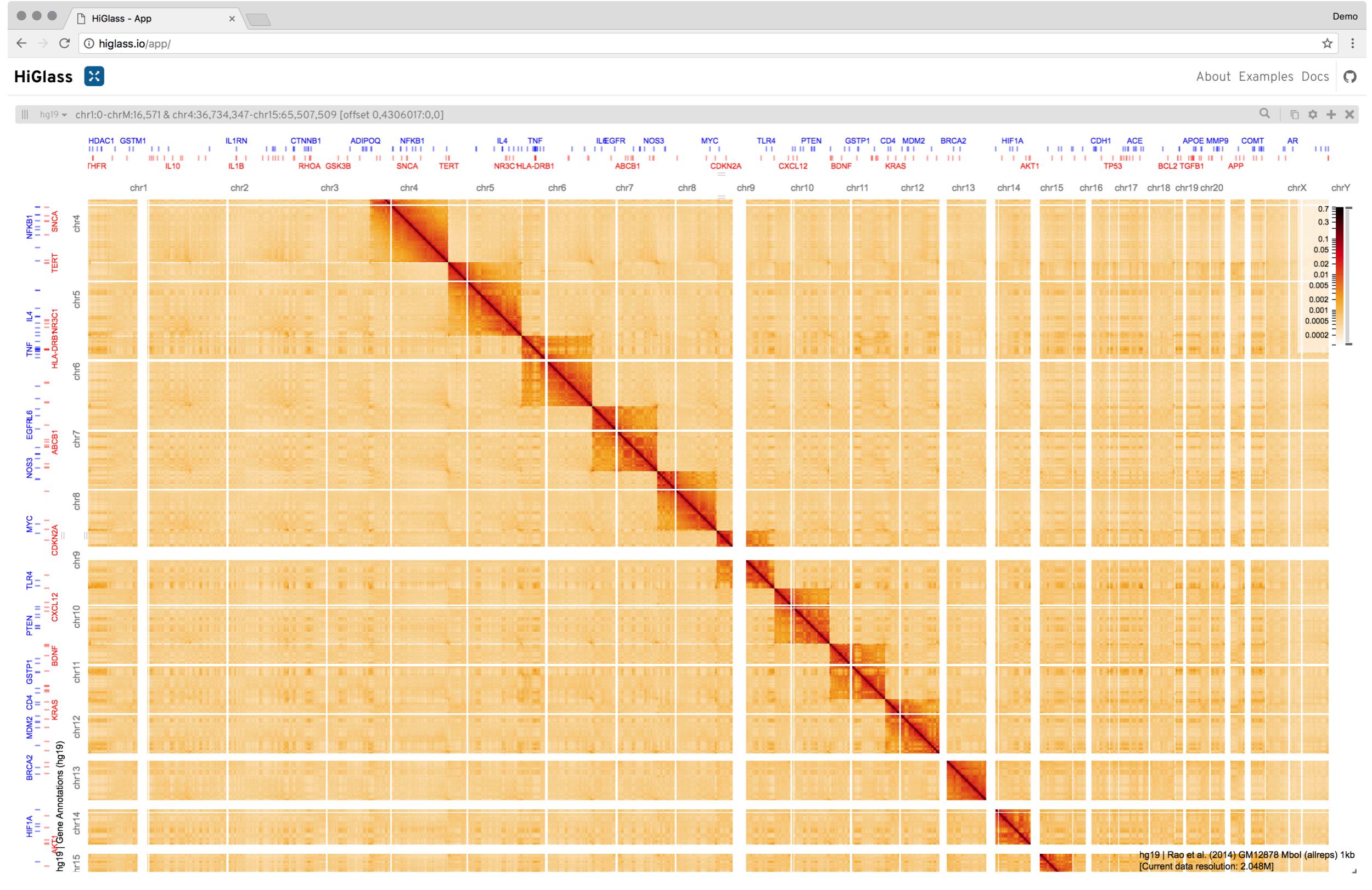
Linearization and Juxtaposition



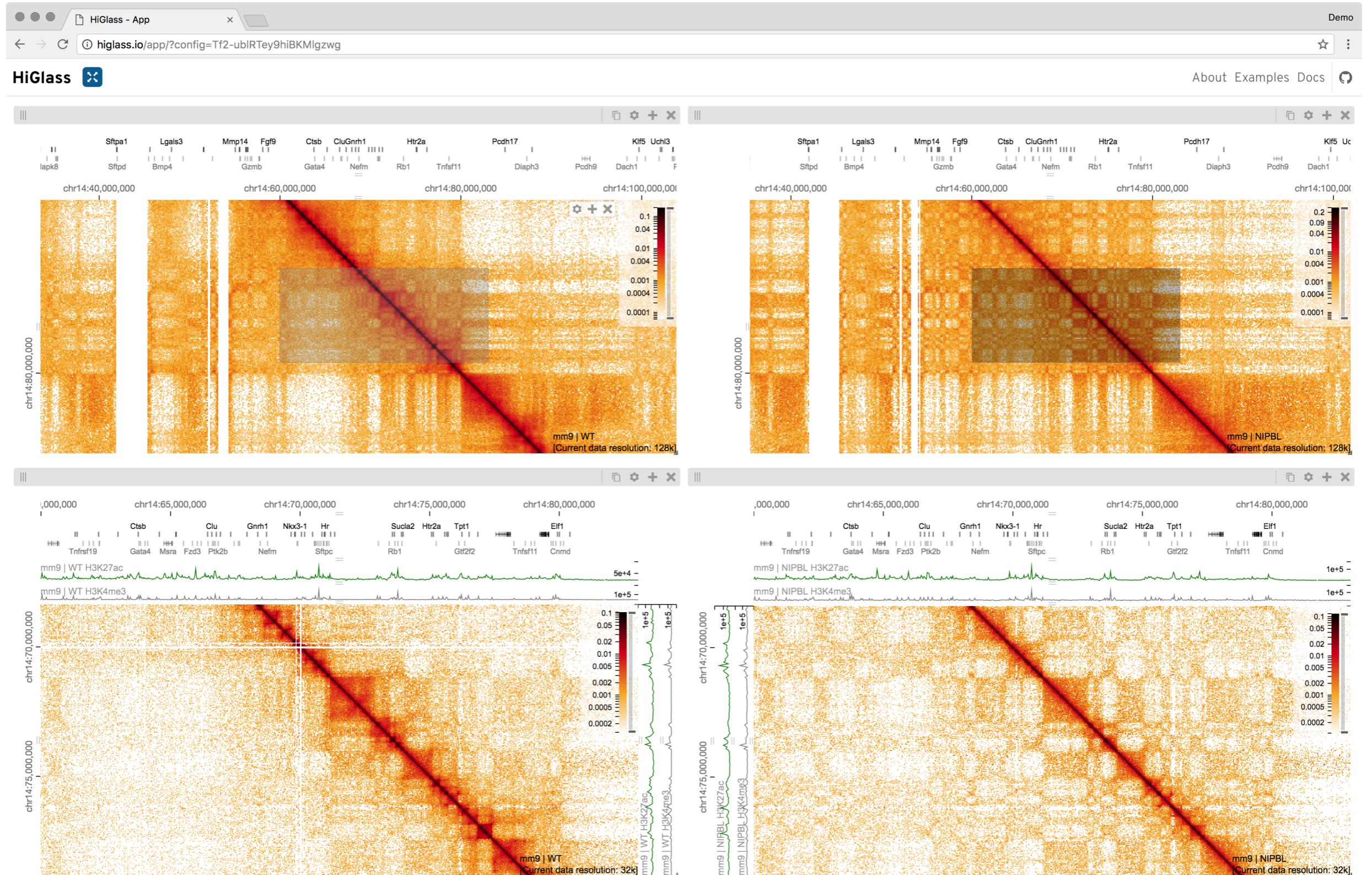


Review of Session 5

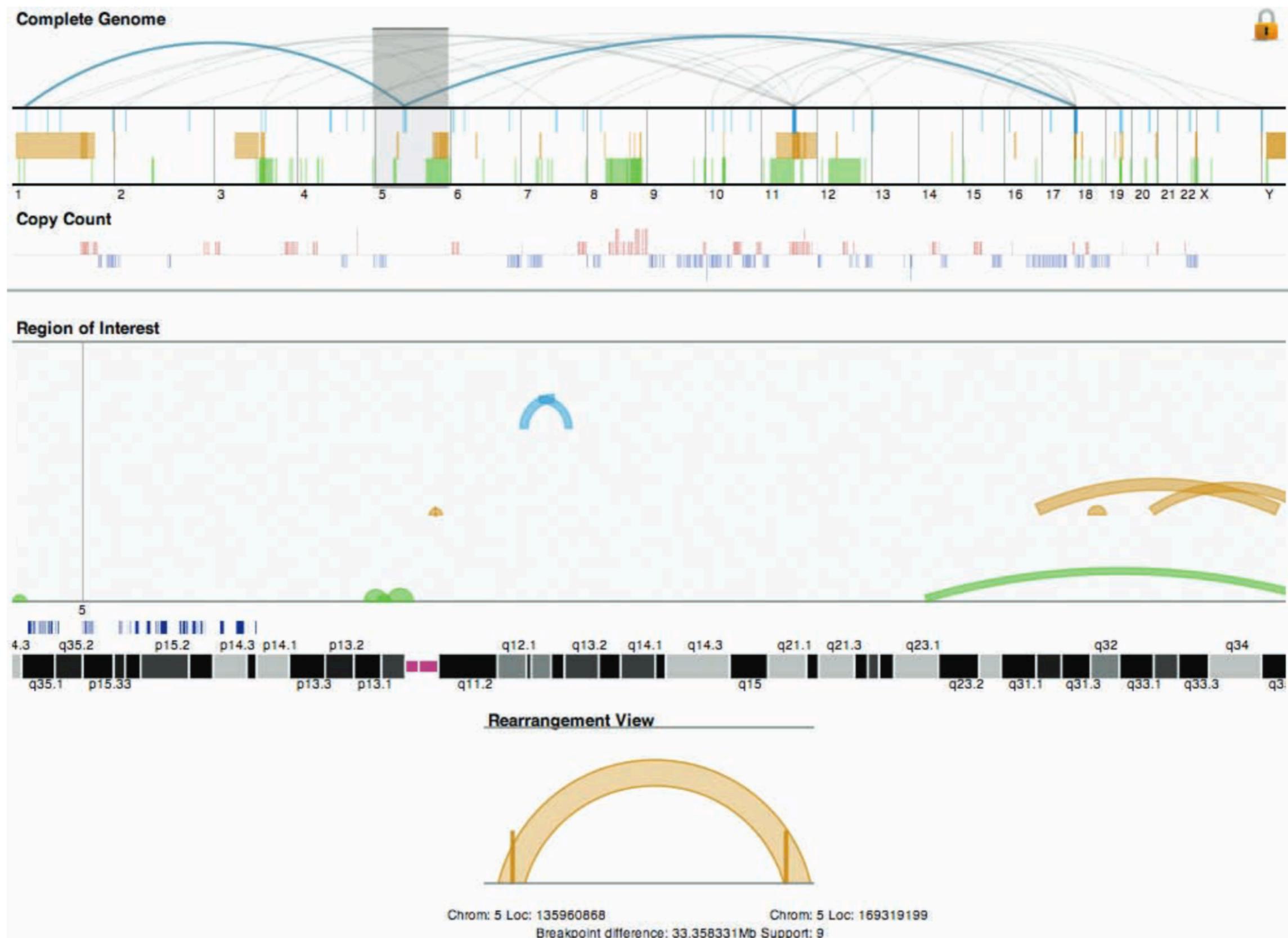
See Peter Kerpeljiev's Slides



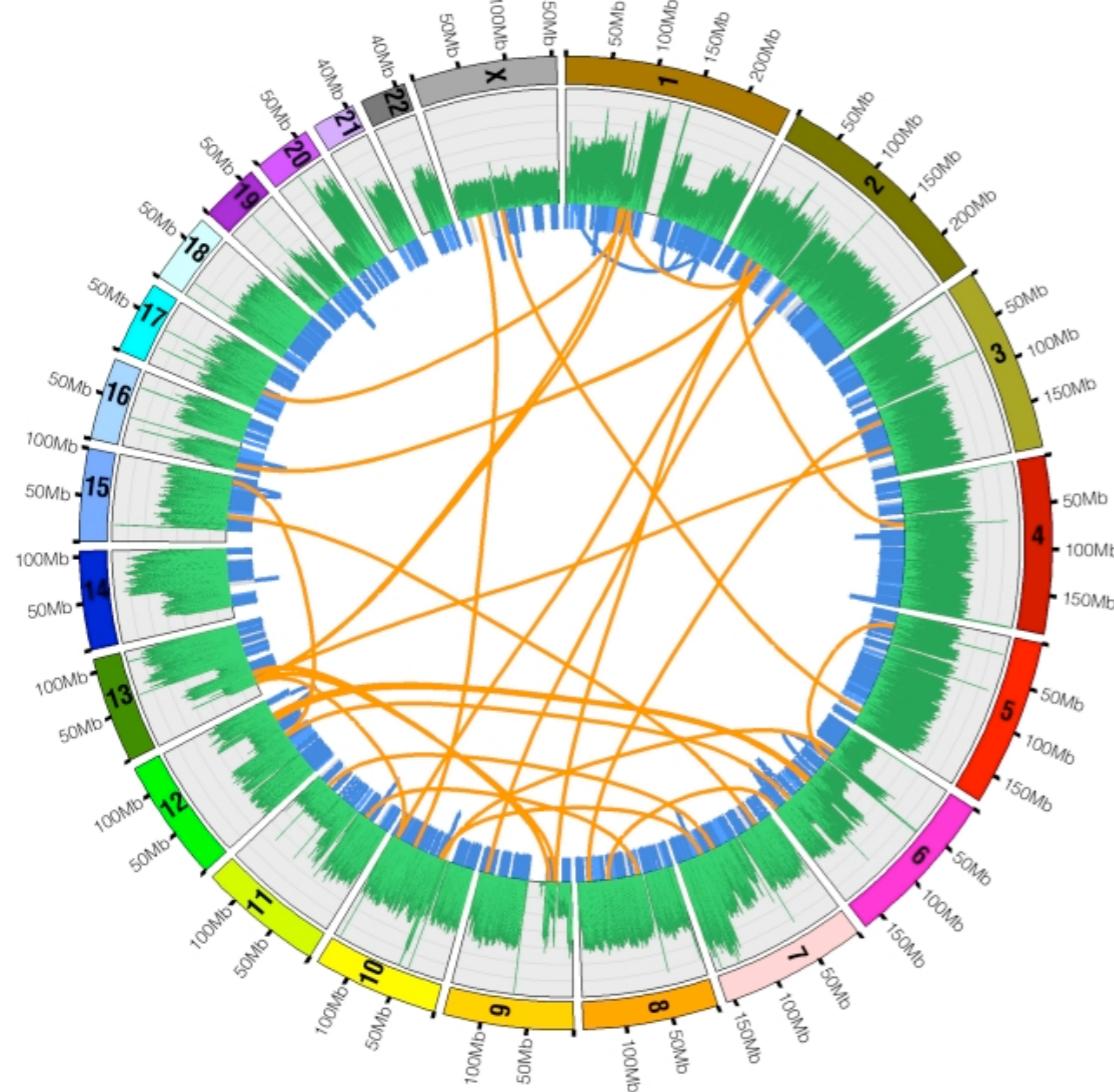
See Peter Kerpédjiev's Slides



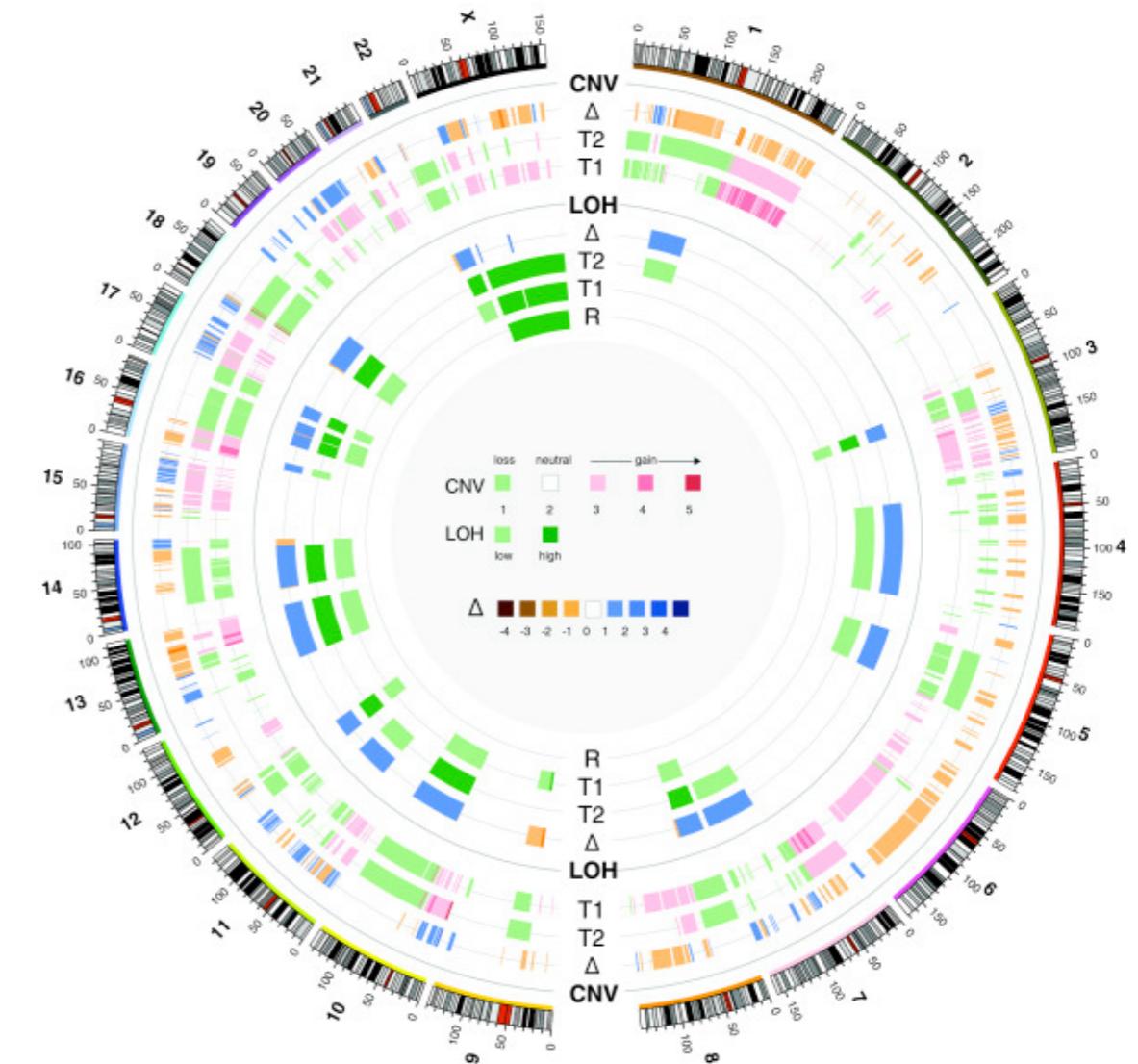
See Gremlin Paper Presentation



Sequences: Circos

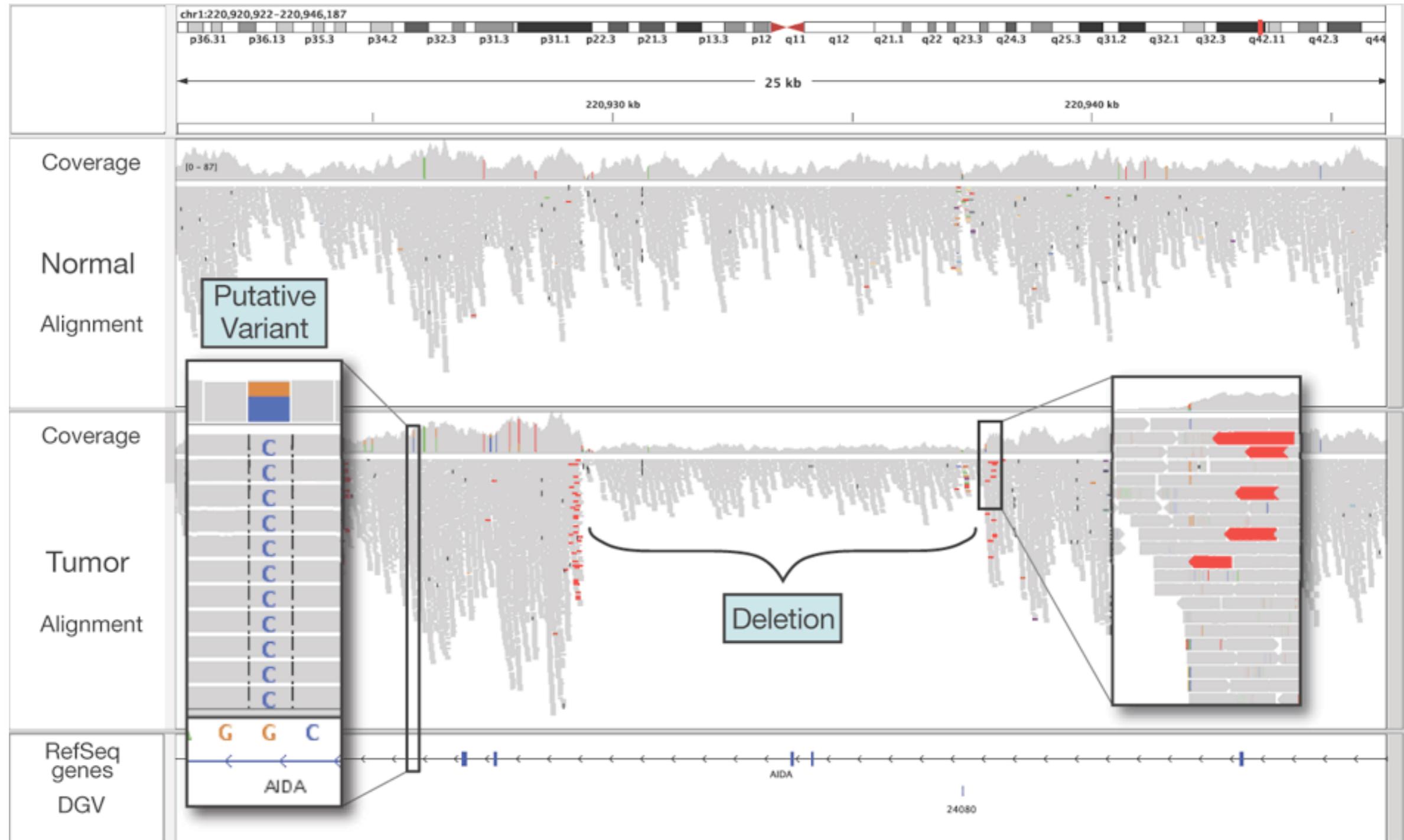


Clark et al. 2009, PLoS Genetics



Jones et al. 2010, Genome Biology

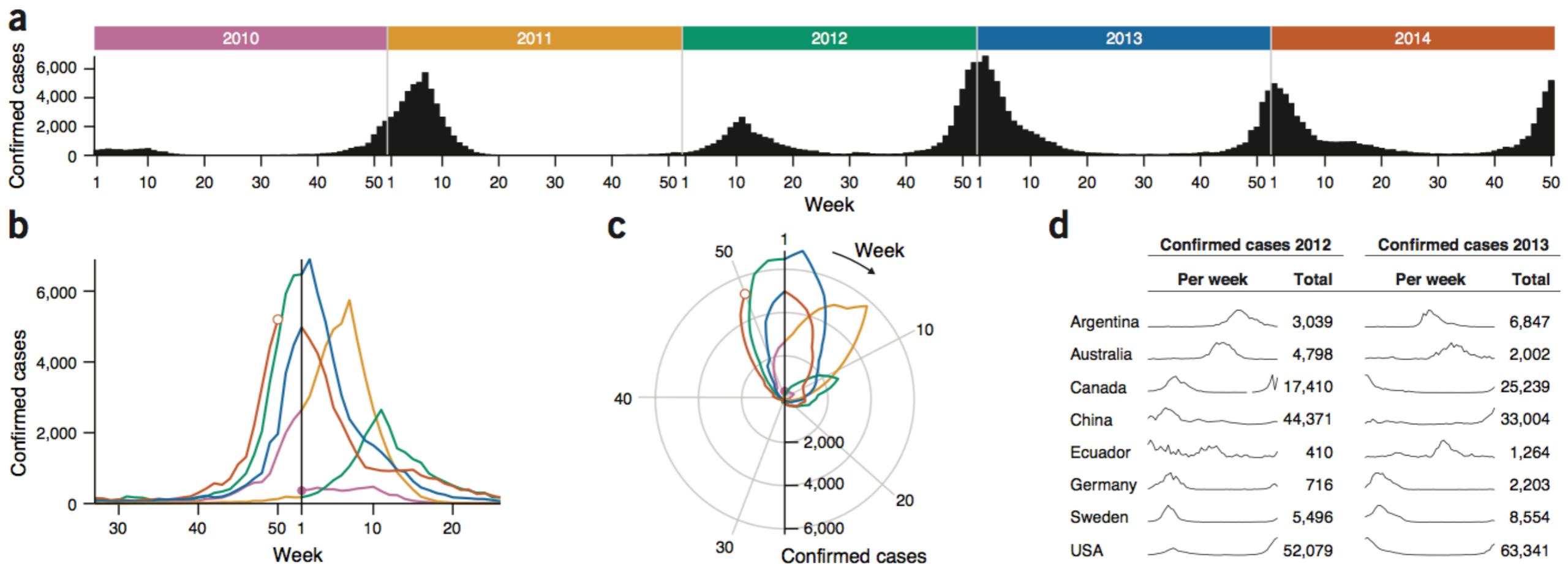
Sequences: Integrative Genomics Viewer (IGV)



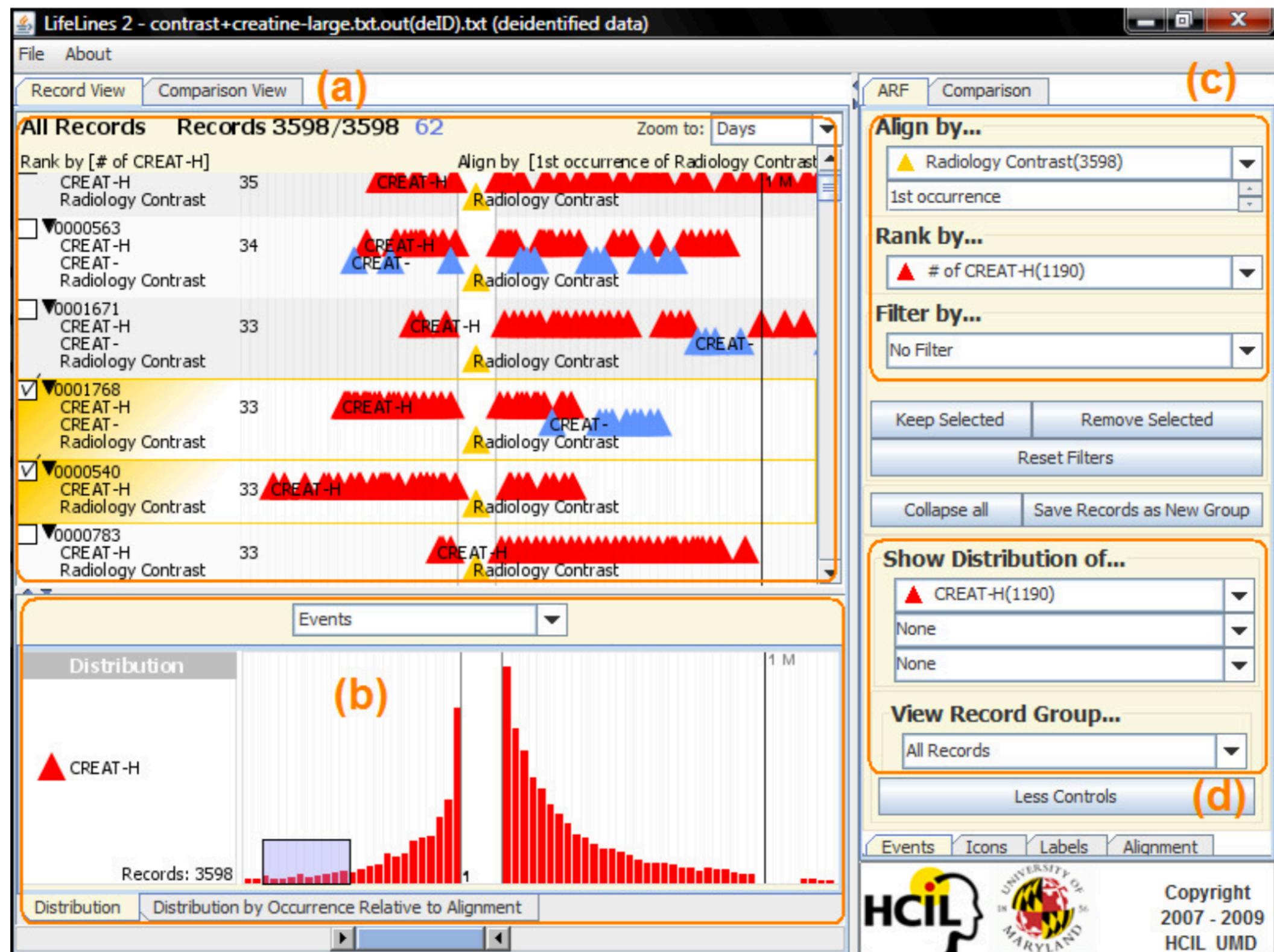
Lecture 6: Learning Goals

- What are the properties of temporal data?
- What are common design approaches for temporal data?
- What are time-oriented visualizations for clinical data?
- What can you do to make your visualizations suitable for presentation?

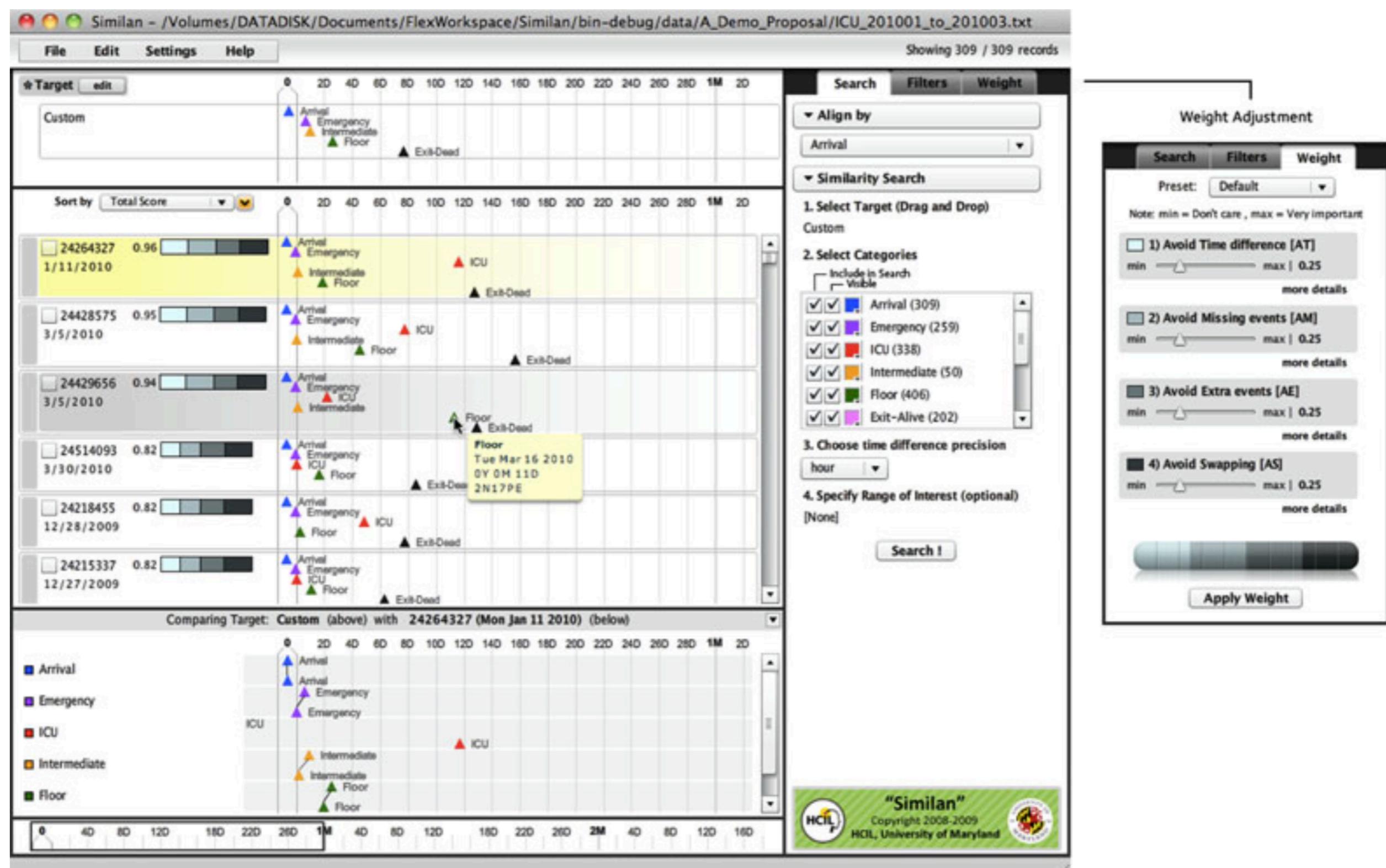
Temporal Data



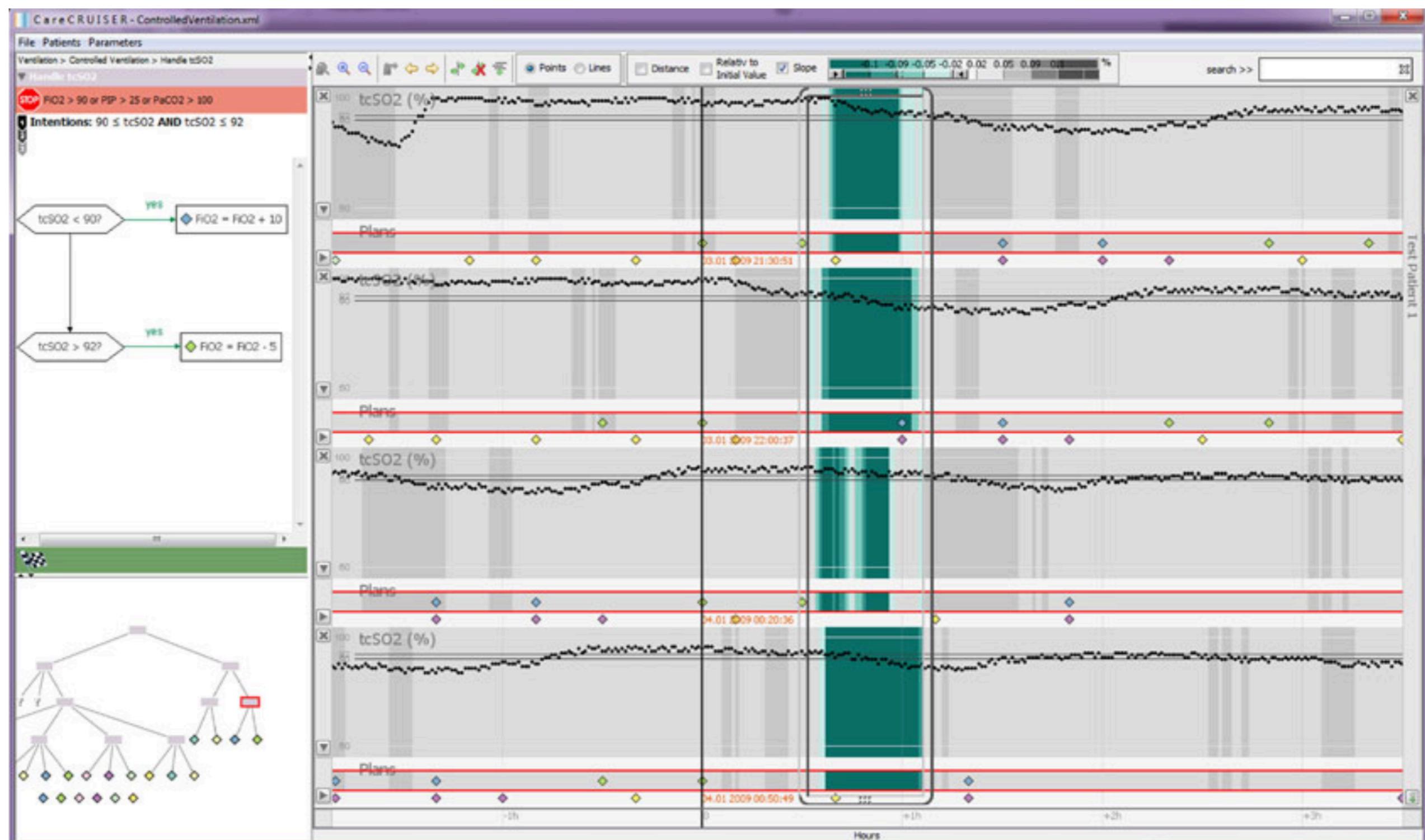
LifeLines 2



Similan



CareCruiser



Temporal Data: Properties

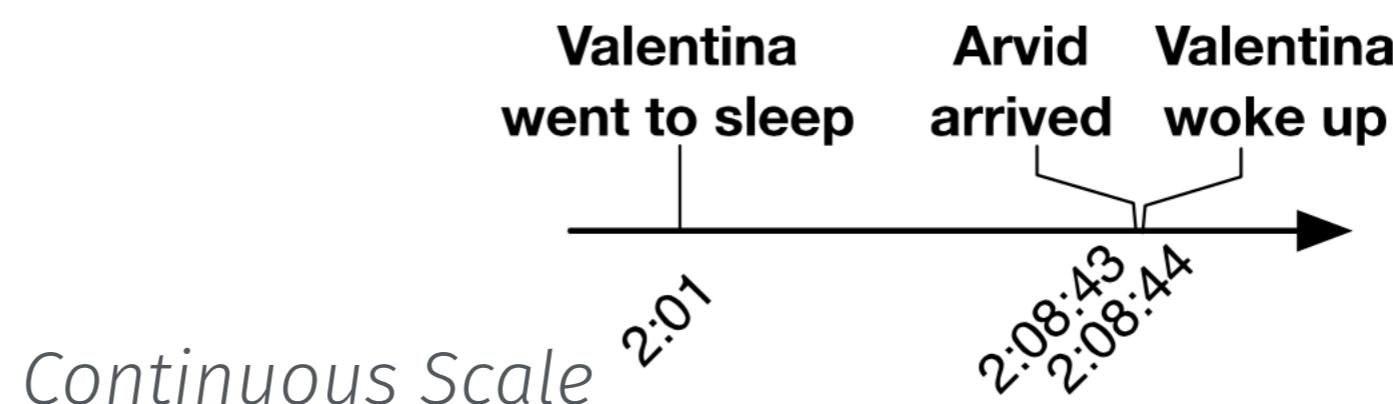
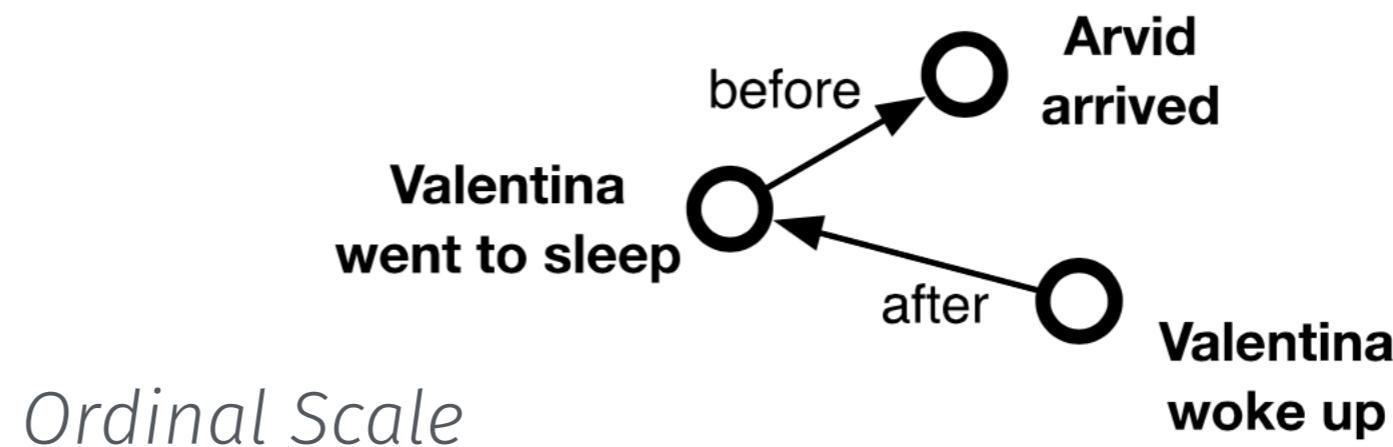
- Scale
- Scope
- Arrangement
- Viewpoint

Scales

Fig. 3.1 Ordinal scale. Only relative order relations are present. At this level it is not possible to discern whether Valentina woke up before or after Arvid arrived.

Fig. 3.2 Discrete scale.
Smallest possible unit is minutes. Although Arvid arrived and Valentina woke up within the same minute, it is not possible to model the exact order of events.

Fig. 3.3 Continuous scale. Between any two points in time, another point in time exists. Here, it is possible to model that Arvid arrived shortly before Valentina woke up.



Scope: Point vs Interval

Fig. 3.7 Time value “August 1, 2008” in a point-based domain. No information is given in between two time points.

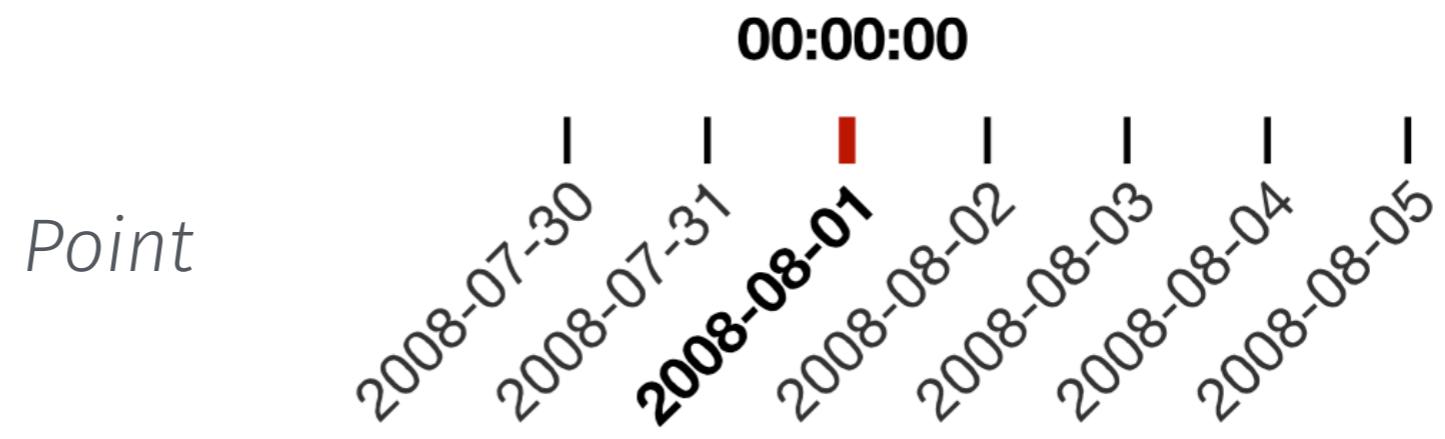
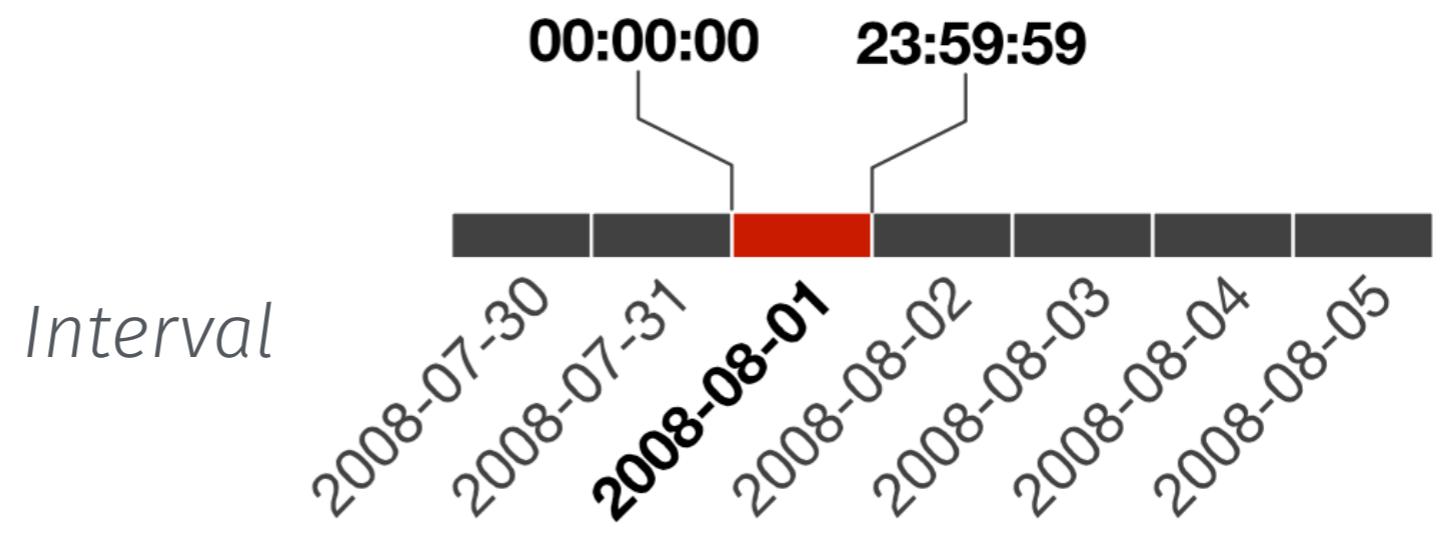


Fig. 3.8 Time value “August 1, 2008” in an interval-based domain. Each element covers a subsection of the time domain greater than zero.

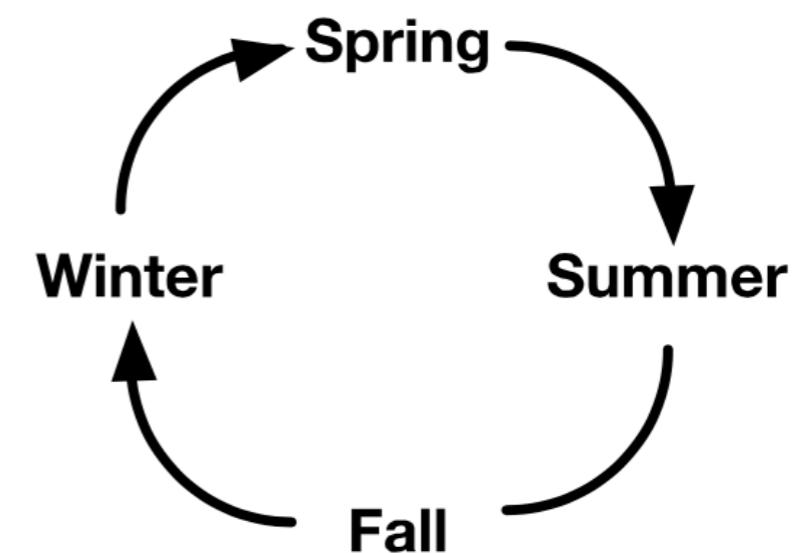


Arrangement: Linear vs Cyclic

Fig. 3.10 Linear time. Time proceeds linearly from past to future.



Fig. 3.11 Cyclic time. Set of recurring time values such as the seasons of the year.

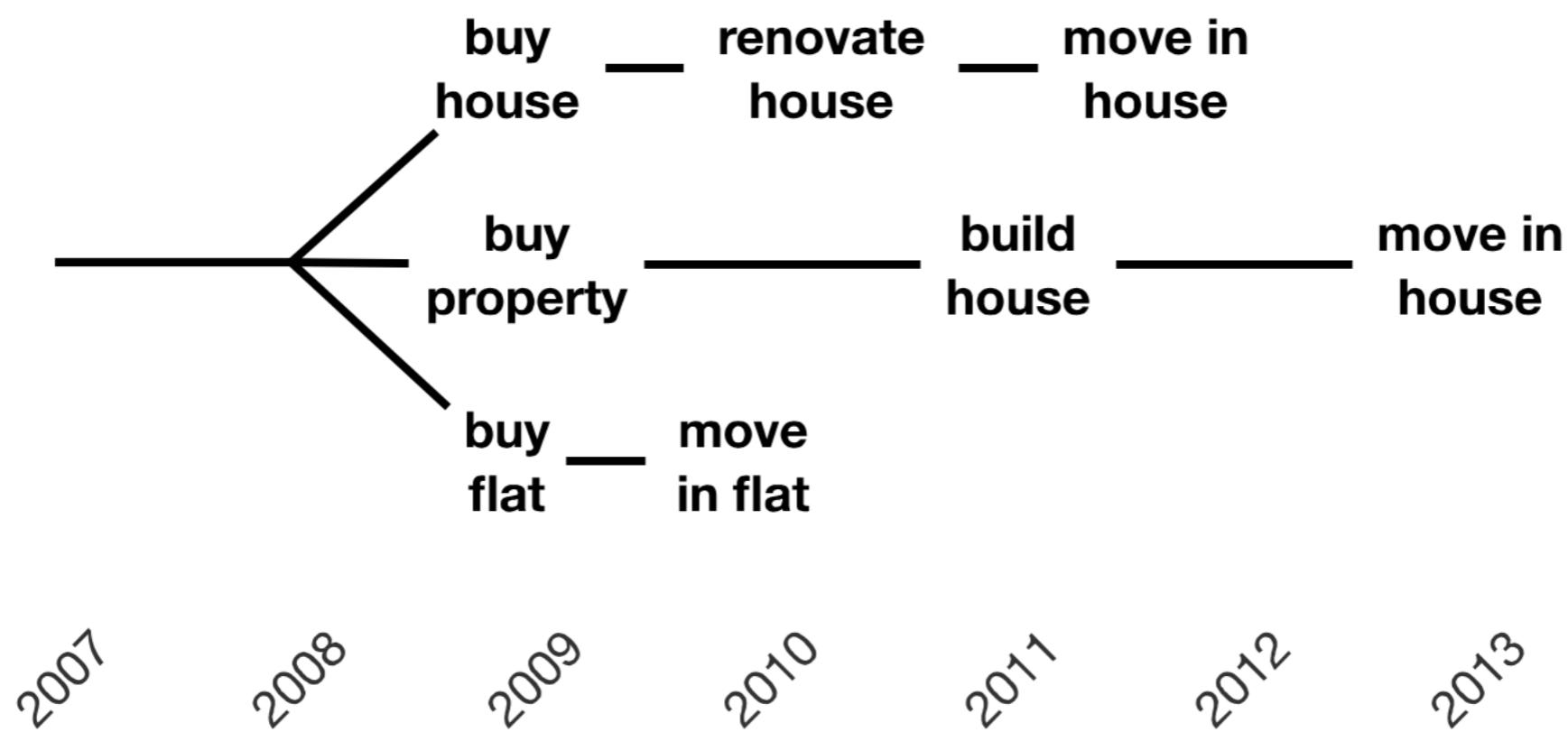


Viewpoint: Ordered

totally ordered: only one event can happen at a time

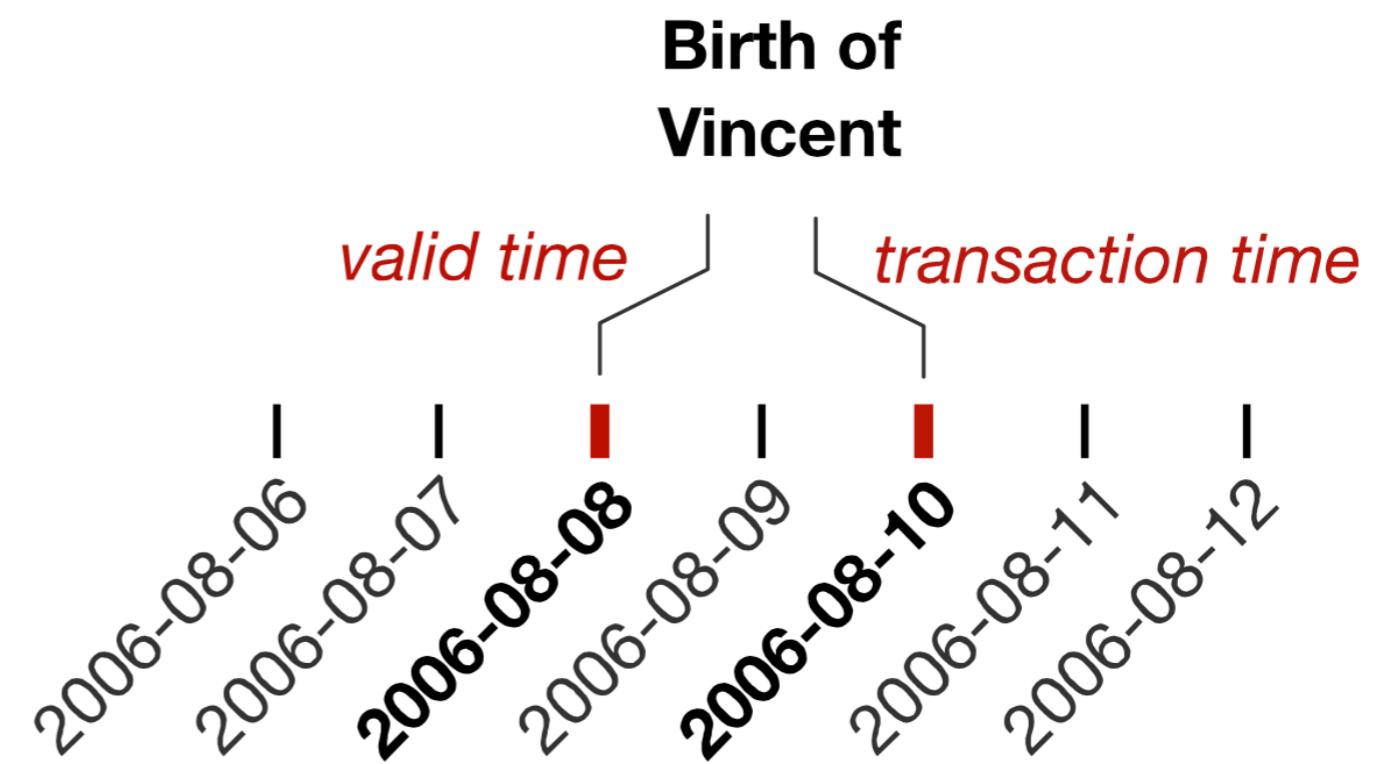
partially ordered: events can overlap

Viewpoint: Branching

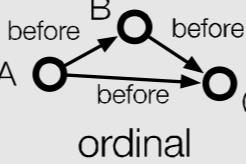
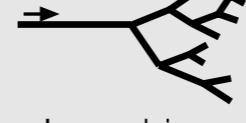
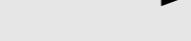


Viewpoint: Multiple Perspectives

Fig. 3.13 Multiple perspectives. Vincent was born on August 8, 2006 (valid time) and this fact was stored in the register of residents two days later on August 10, 2006 (transaction time).



Temporal Data: Properties

scale	 ordinal	 discrete	 continuous
scope		point-based	 interval-based
arrangement		linear	 cyclic
viewpoint		ordered	 branching
Abstractions			
granularity & calendars		 single	 multiple
time primitives		instant	 interval
determinacy		determinate	 indeterminate

Temporal Data: Design Space

What?

time

scale, scope, arrangement, viewpoint
granularity & calendars, time primitives,
determinacy
see Chapter 3

data

scale, frame of reference, kind of data,
number of variables
see Chapter 3

time & data

internal time, external time
see Chapter 3

Why?

1st level



individual values



sets

2nd level



lookup

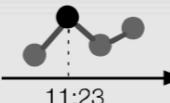


comparison

3rd level



identification



localization

How?

mapping



static



dynamic

dimensionality



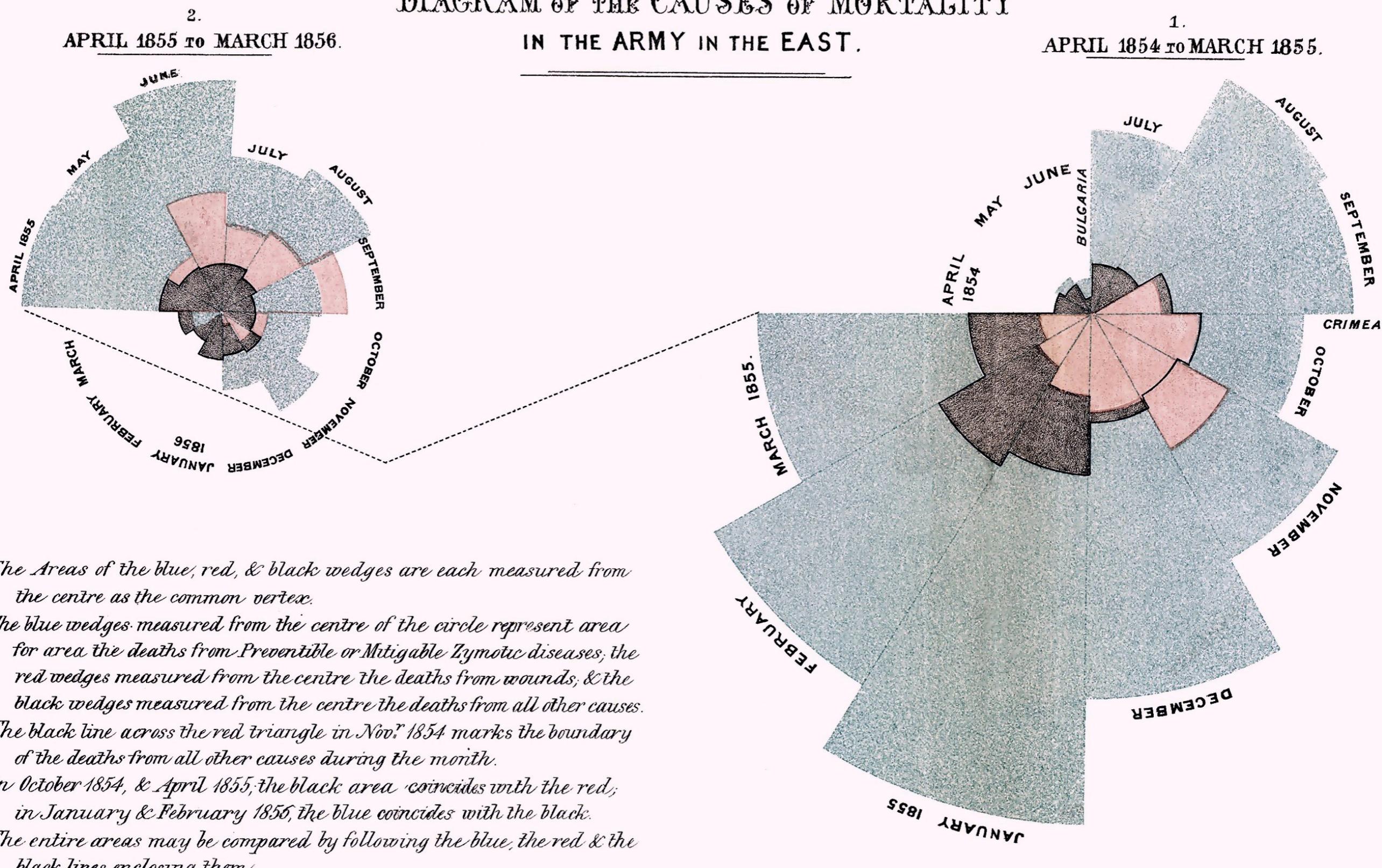
2D



3D

Nightingale: Rose Charts (1858)

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

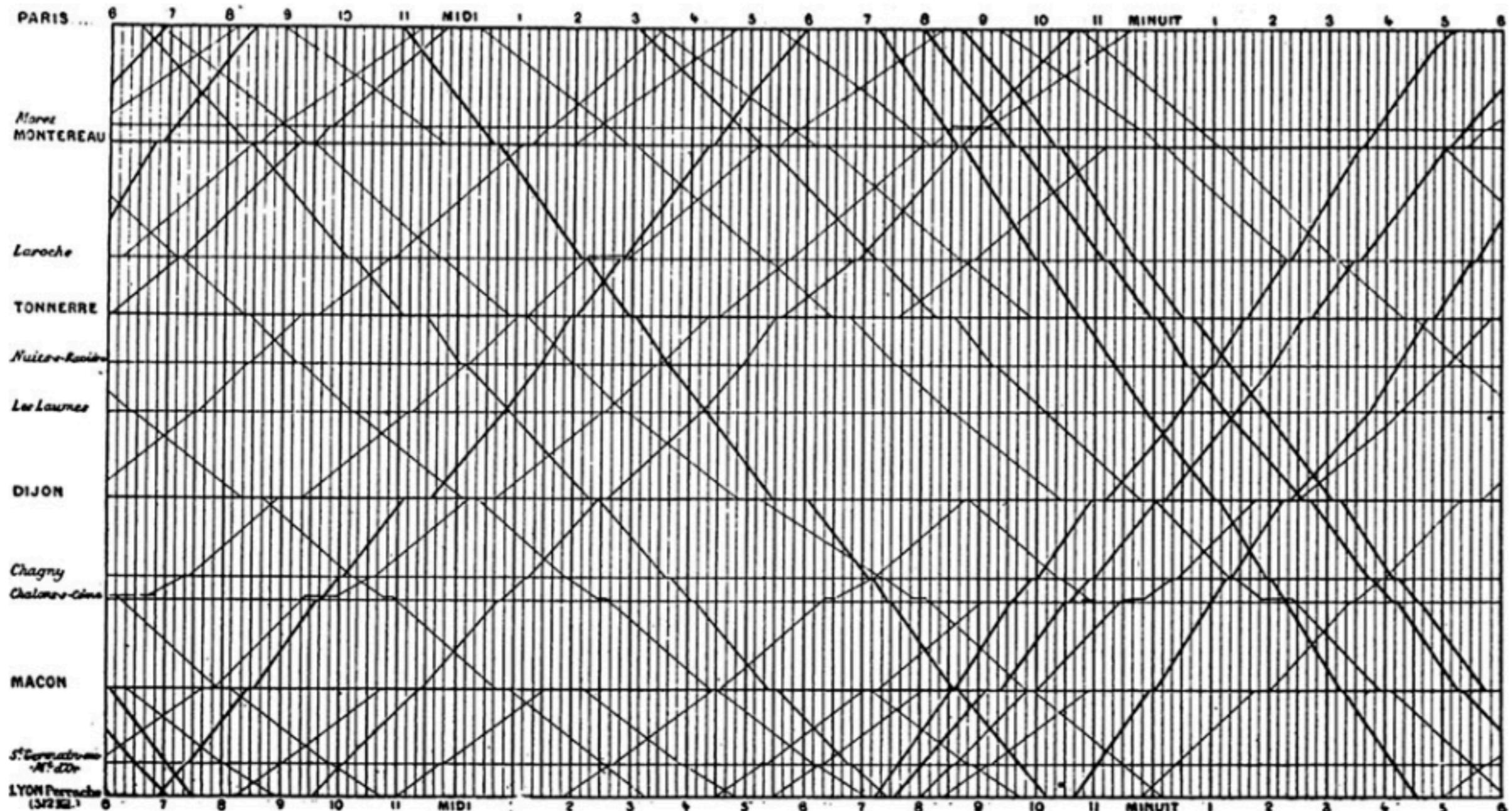
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov.^r 1854 marks the boundary of the deaths from all other causes during the month.

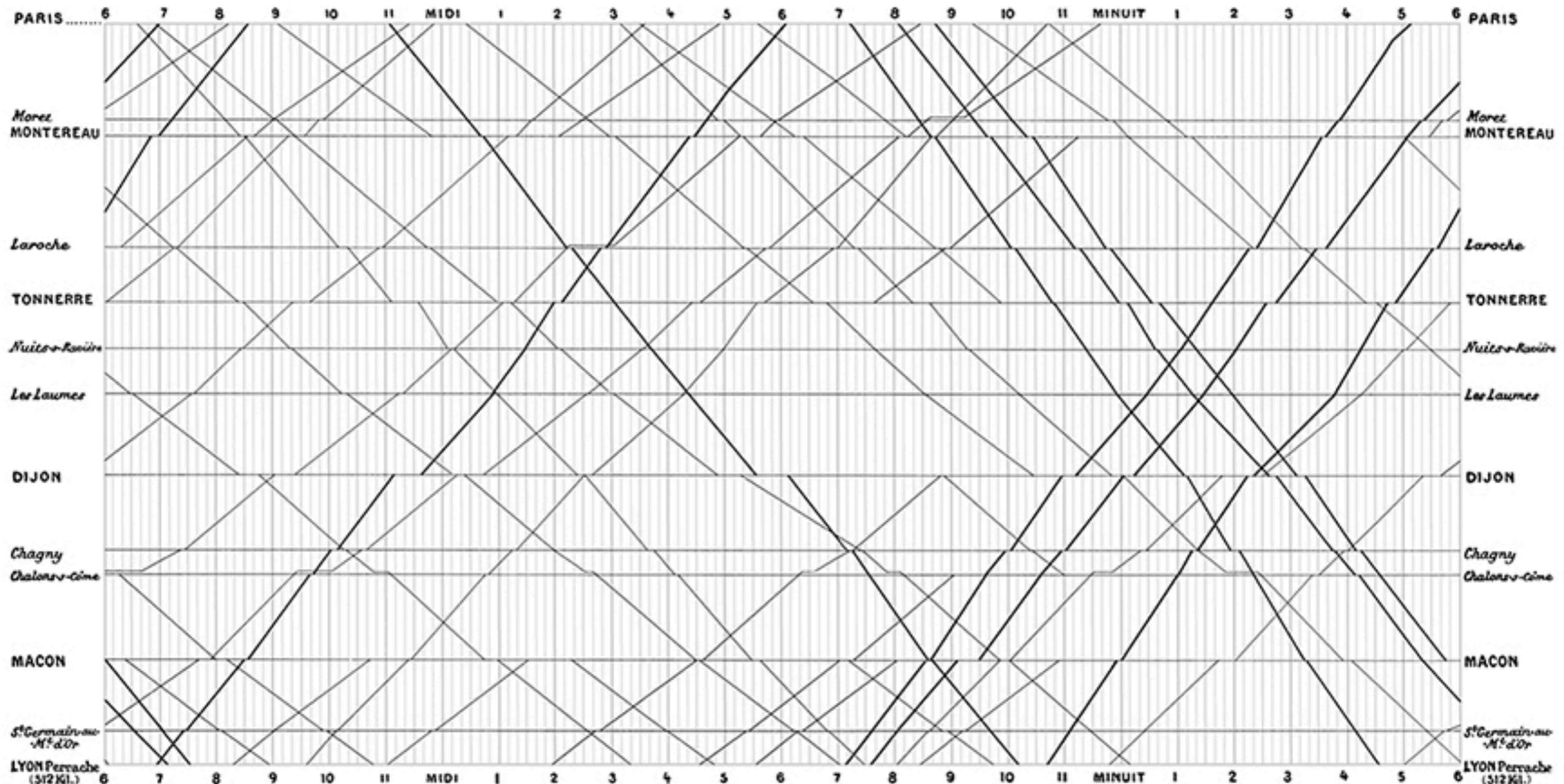
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

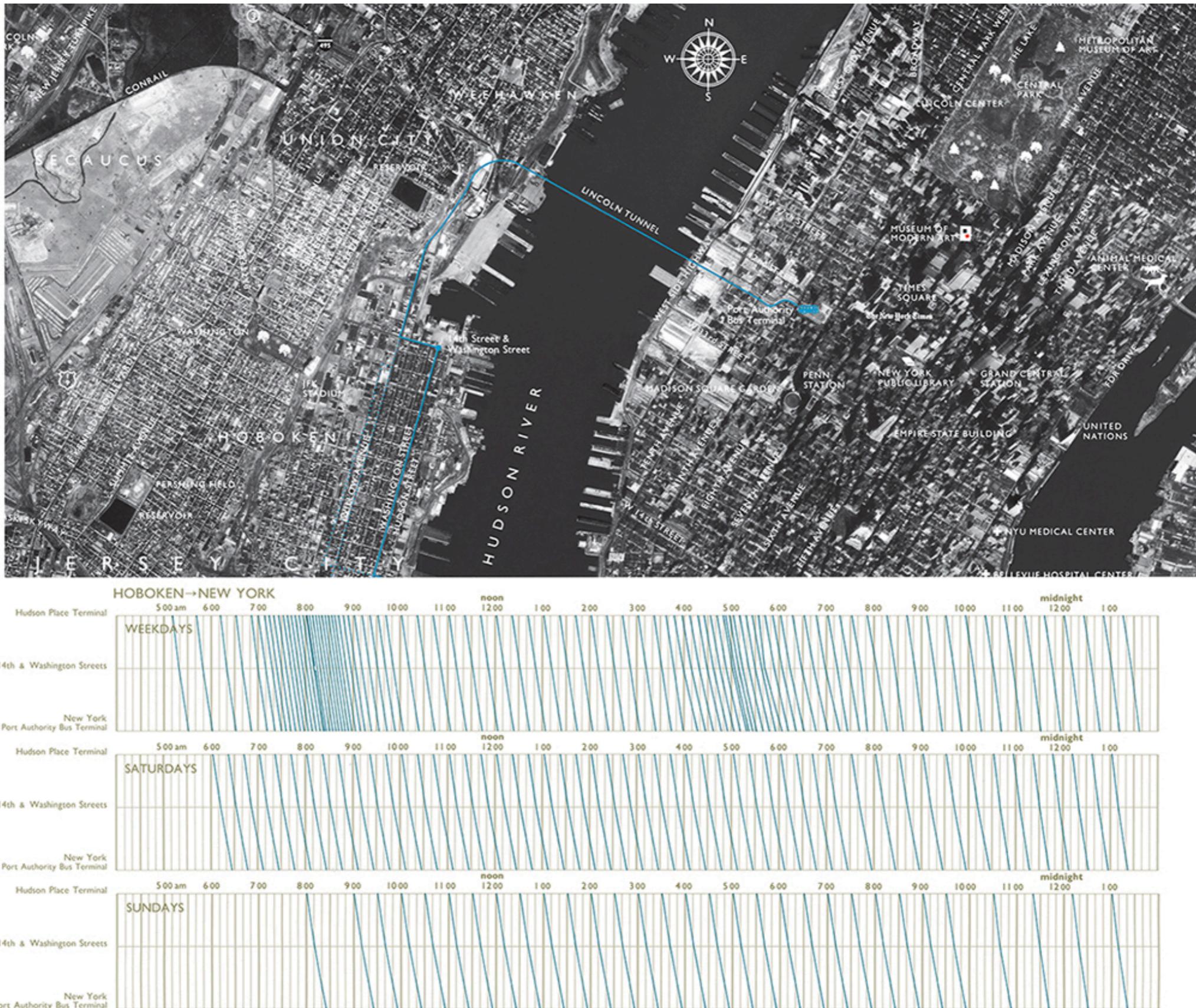
Marey: Train Schedule (1875)



Marey: Train Schedule (1875) - Redesigned by Tufte (1983)

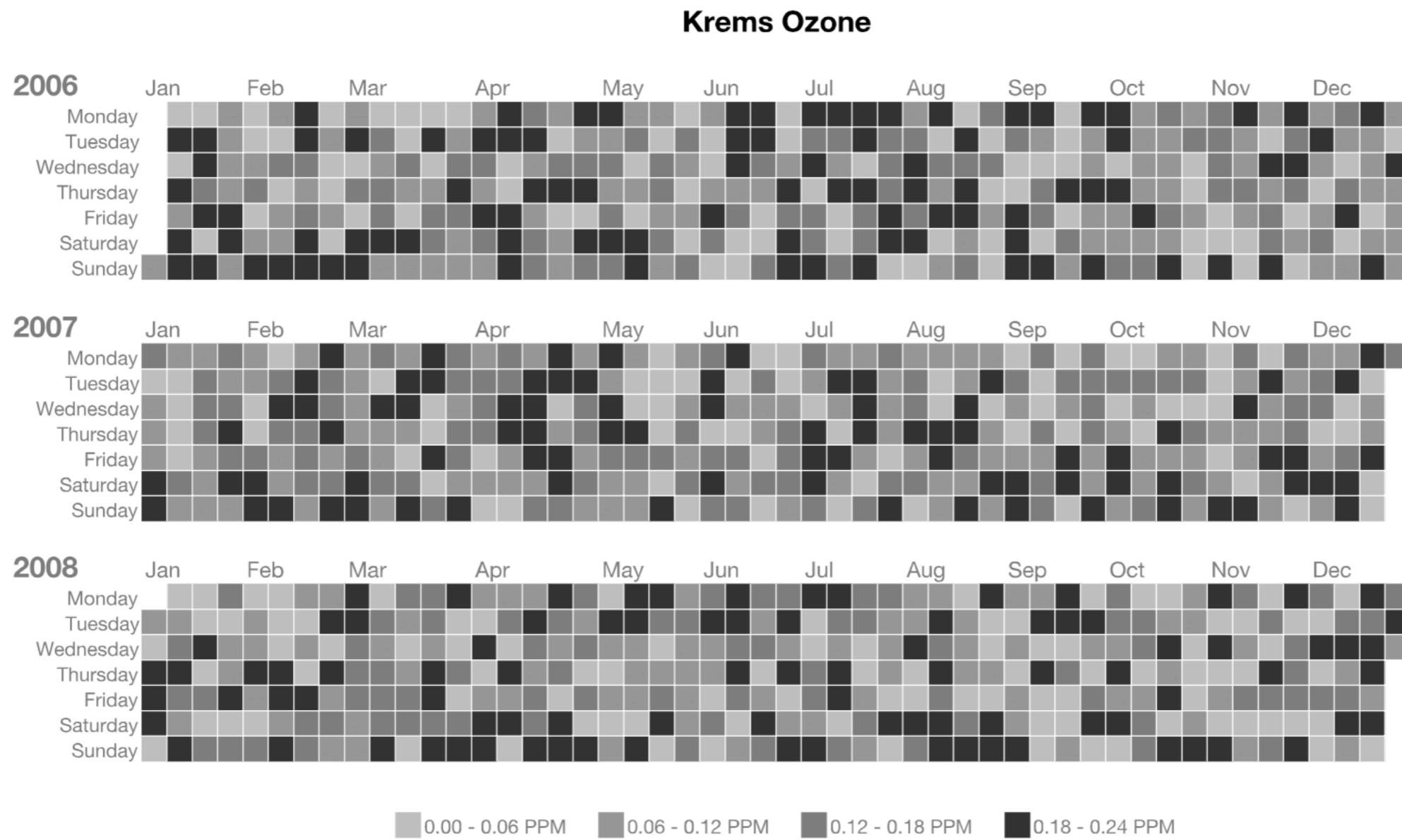


Tufte: Bus Schedule (1990)



from: Tufte, Envisioning Information (1990)

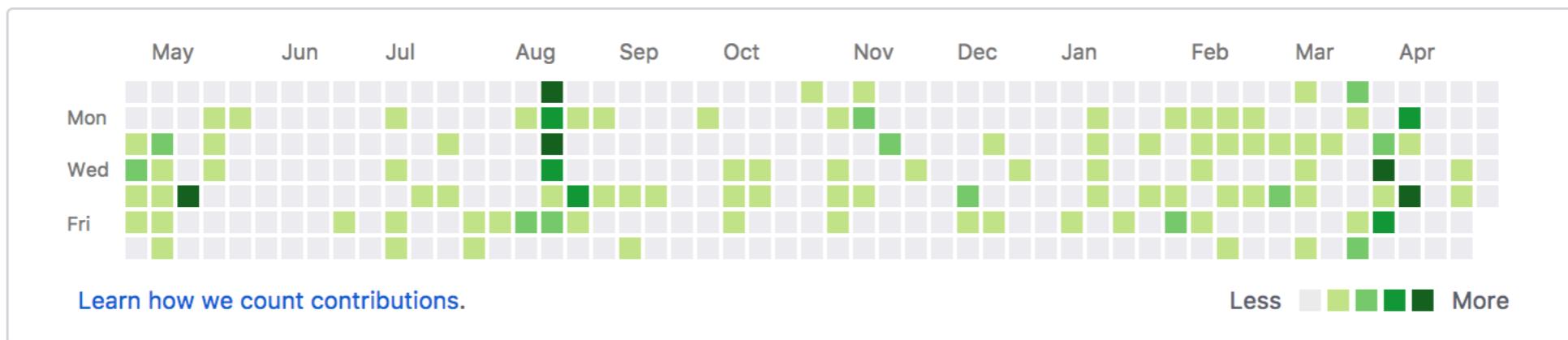
Tile Maps



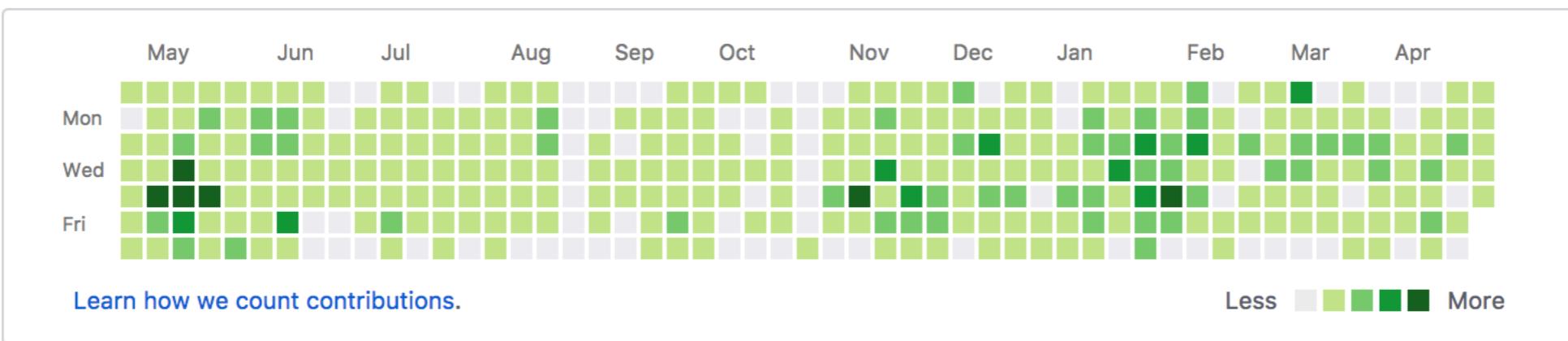
from: Aigner et al., Visualization of Time-oriented Data (2011)

Tile Maps

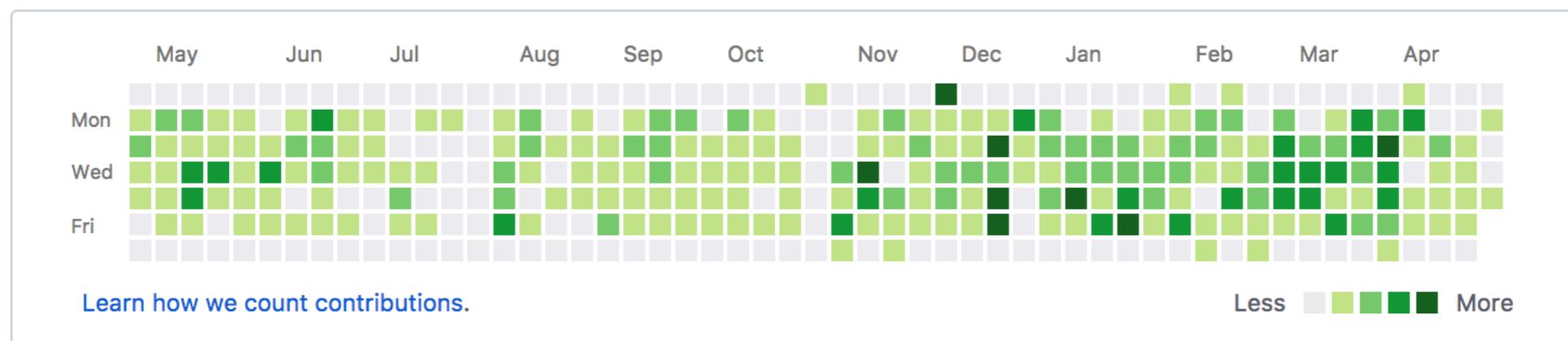
492 contributions in the last year



3,259 contributions in the last year



1,884 contributions in the last year



Calendar-based Views

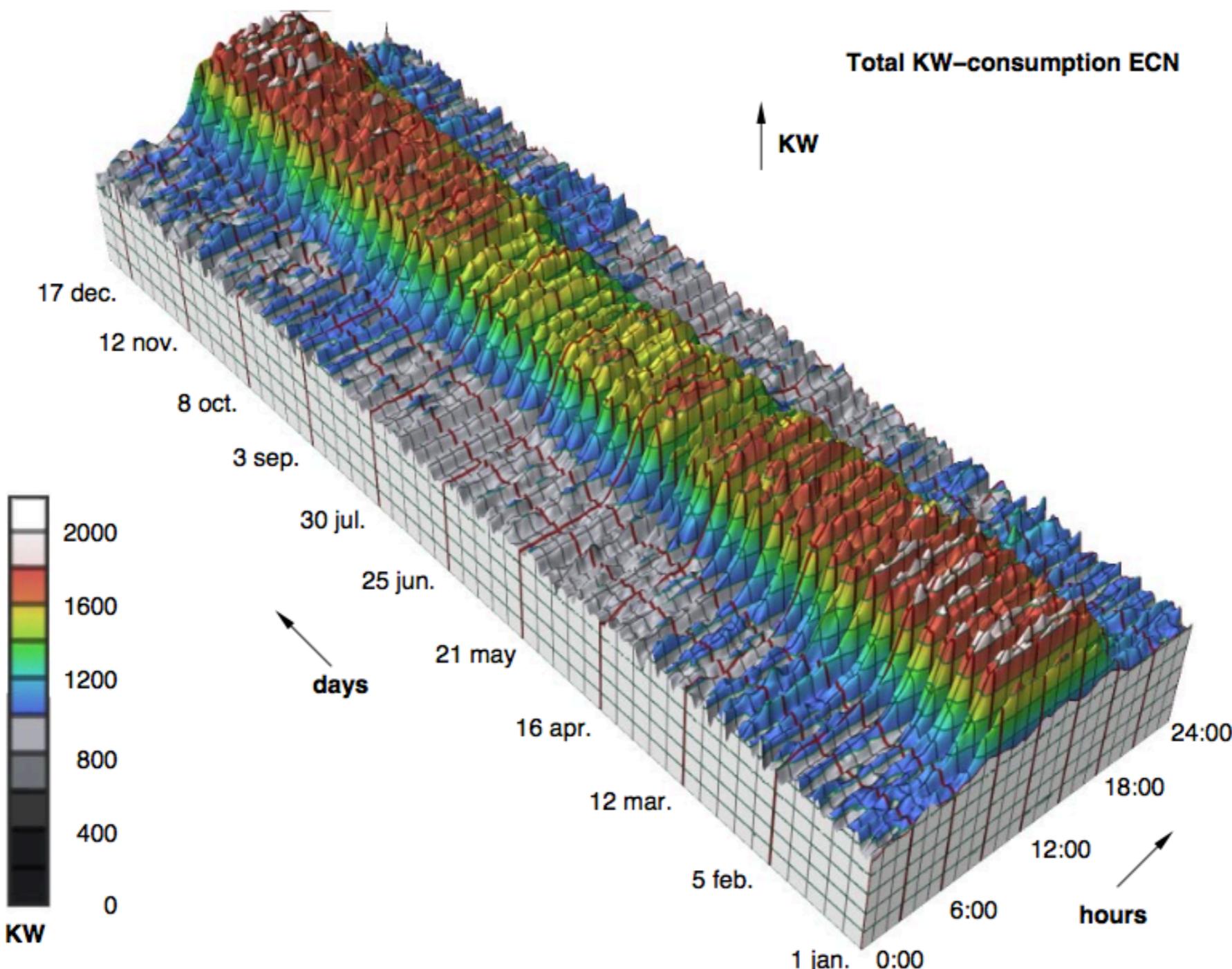


Figure 1. Power demand by ECN, displayed as a function of hours and days

Calendar-based Views

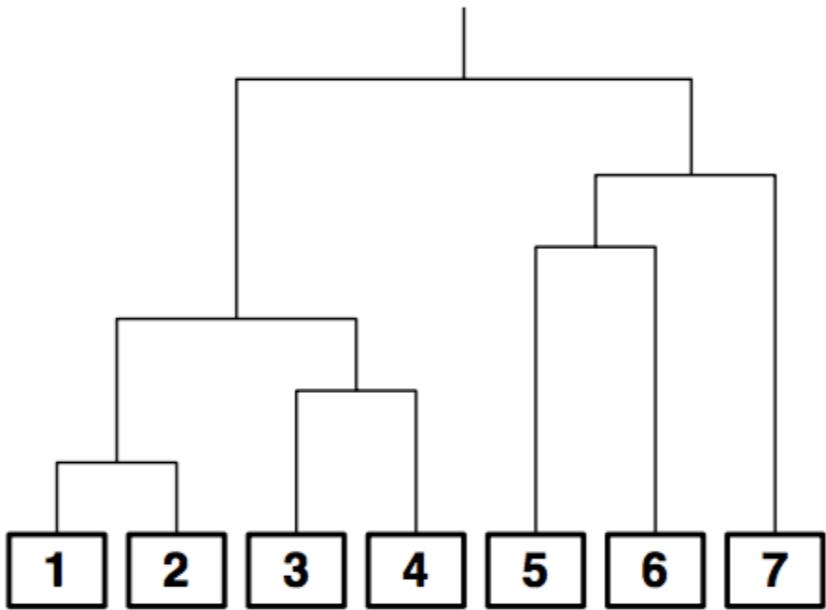


Figure 2. Dendrogram

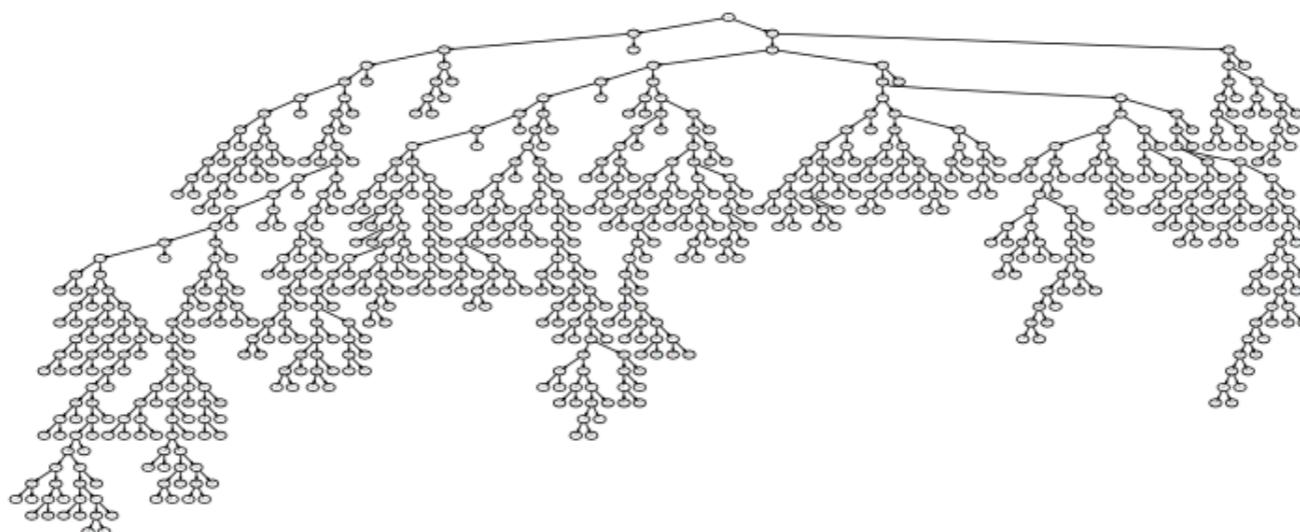


Figure 3. Full clustering tree

Calendar-based Views

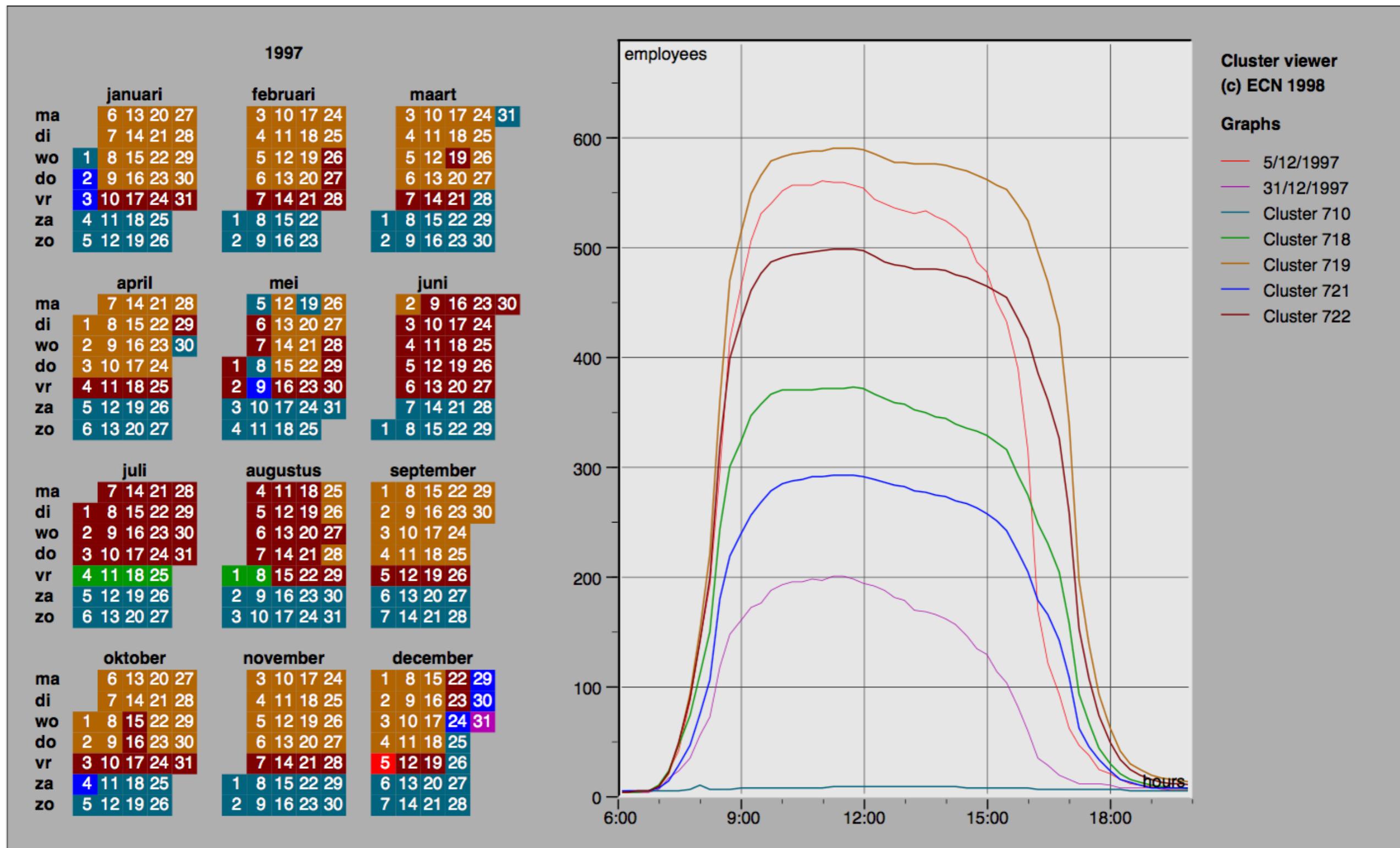
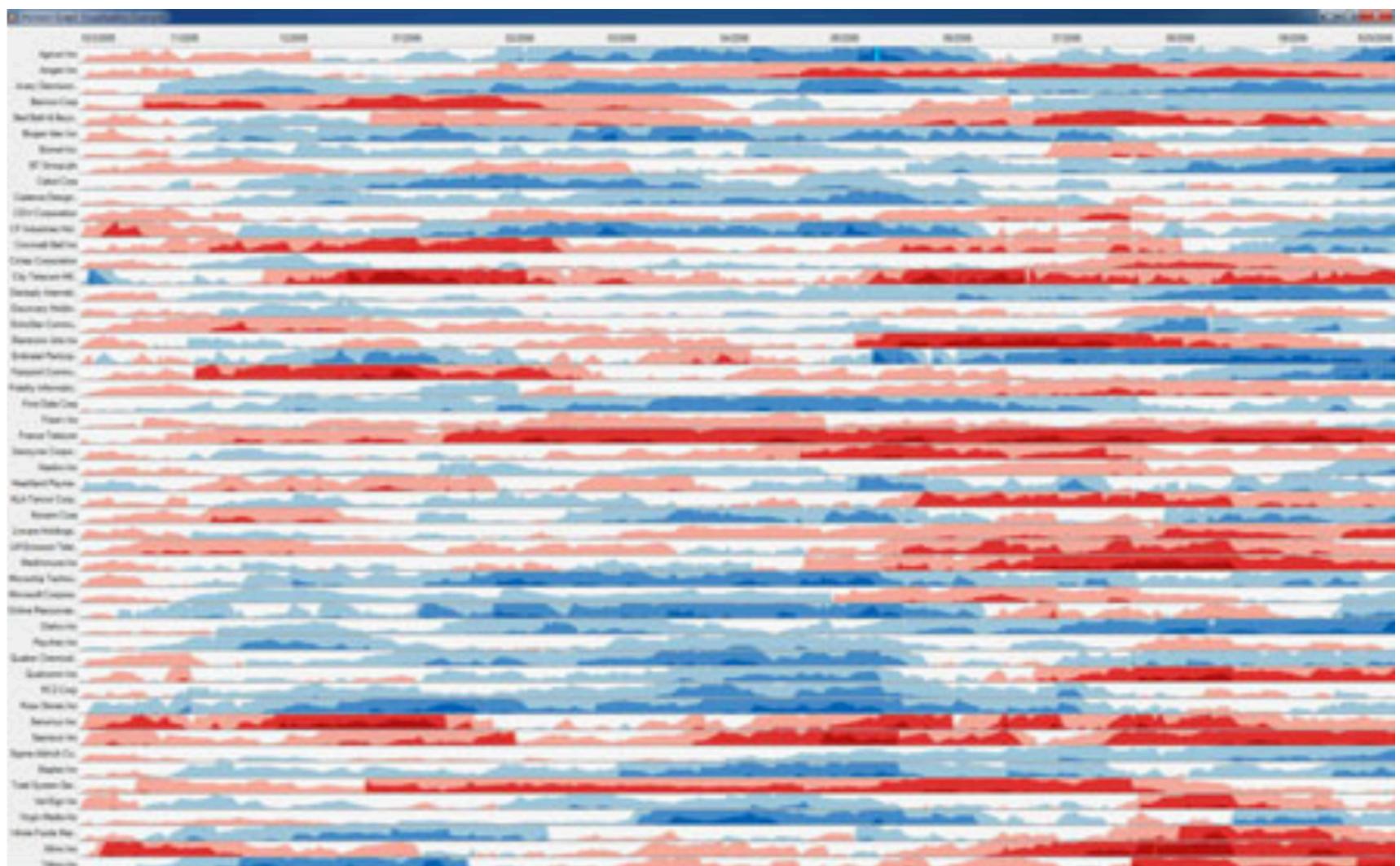
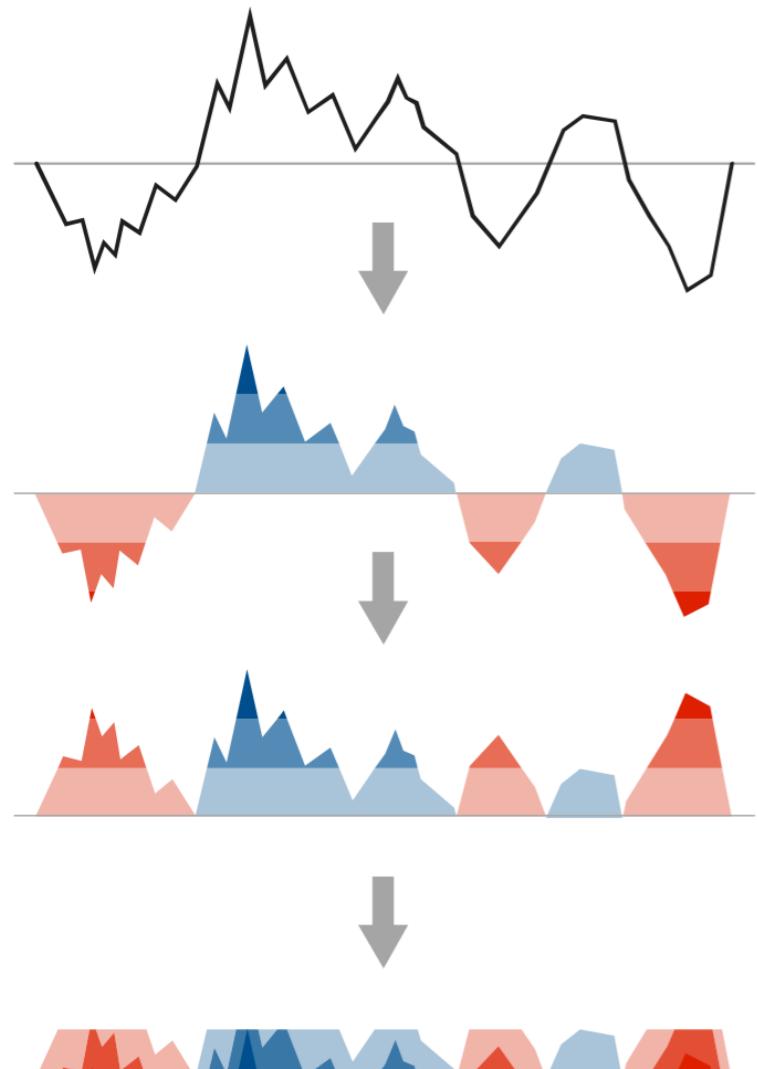


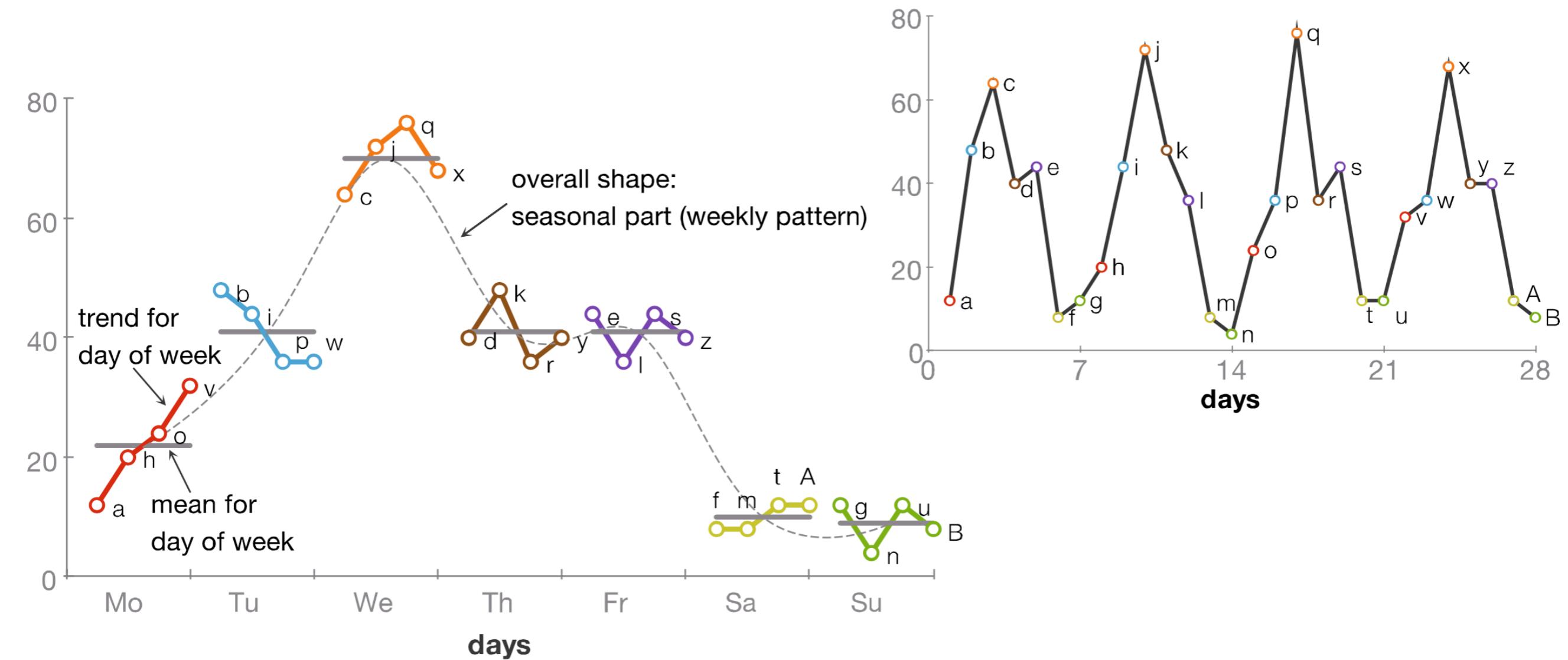
Figure 4. Calendar view of the number of employees

Horizon Charts

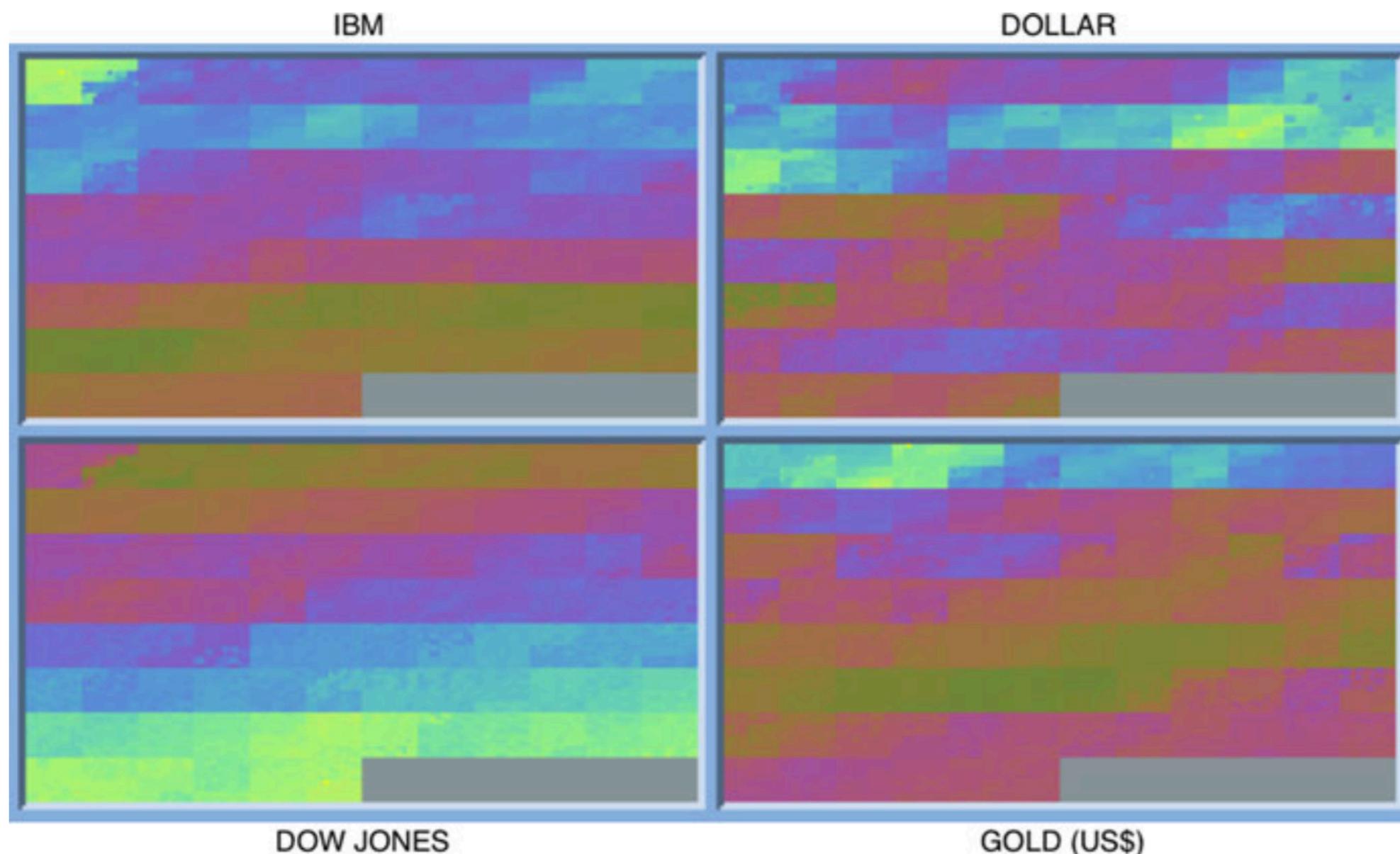


from: Aigner et al., Visualization of Time-oriented Data (2011)

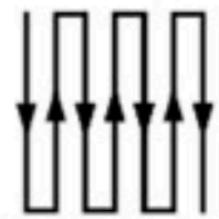
Cycle Plot



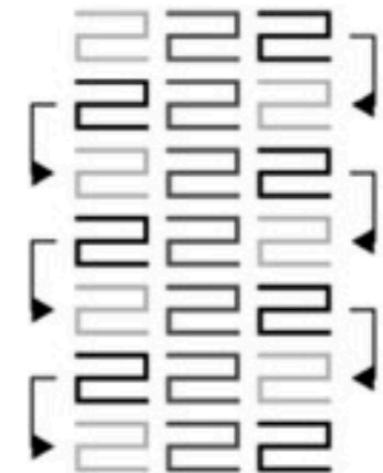
Pixel Maps



a. left-right

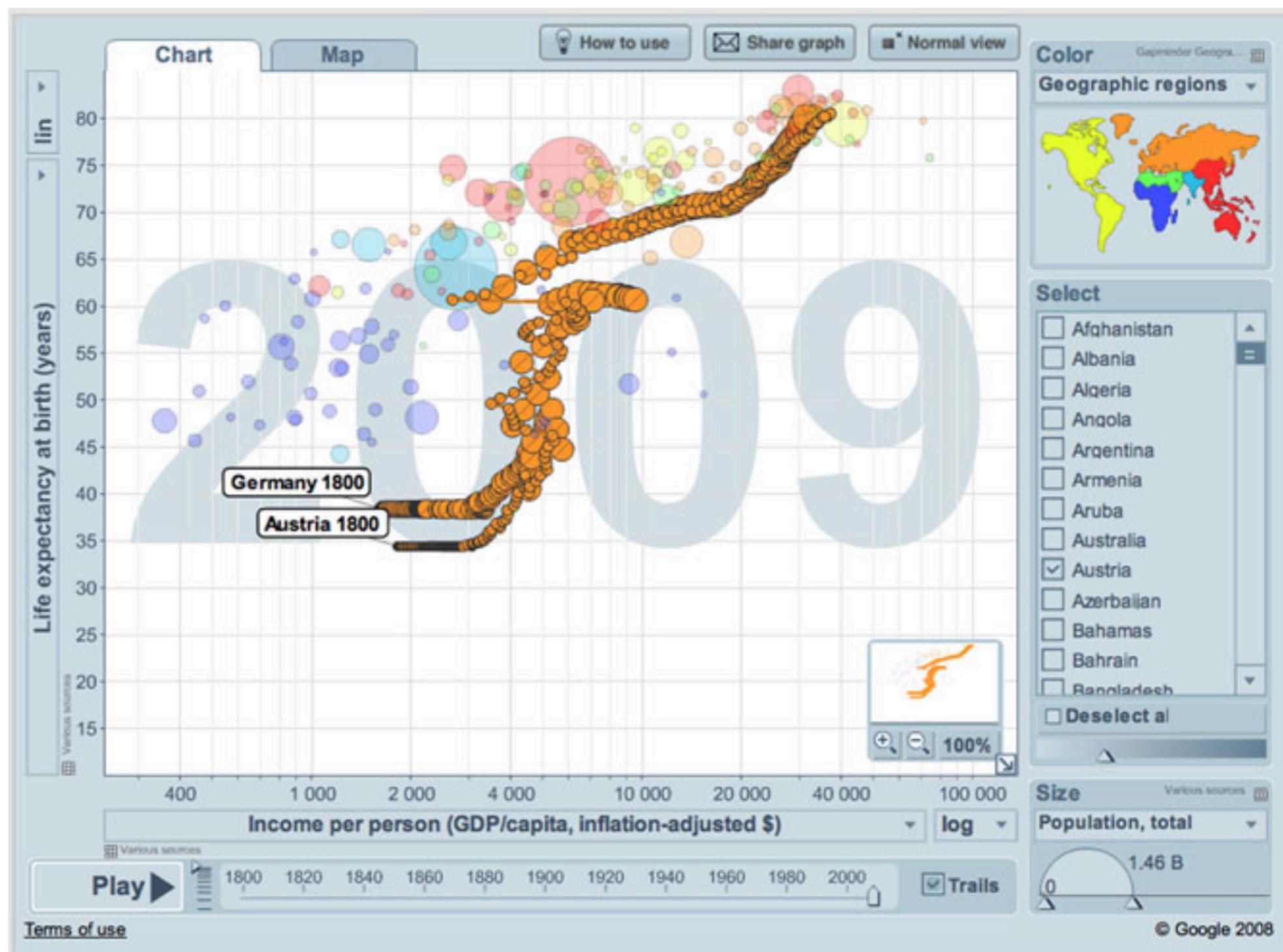


b. top-down



c. back-and-forth loop

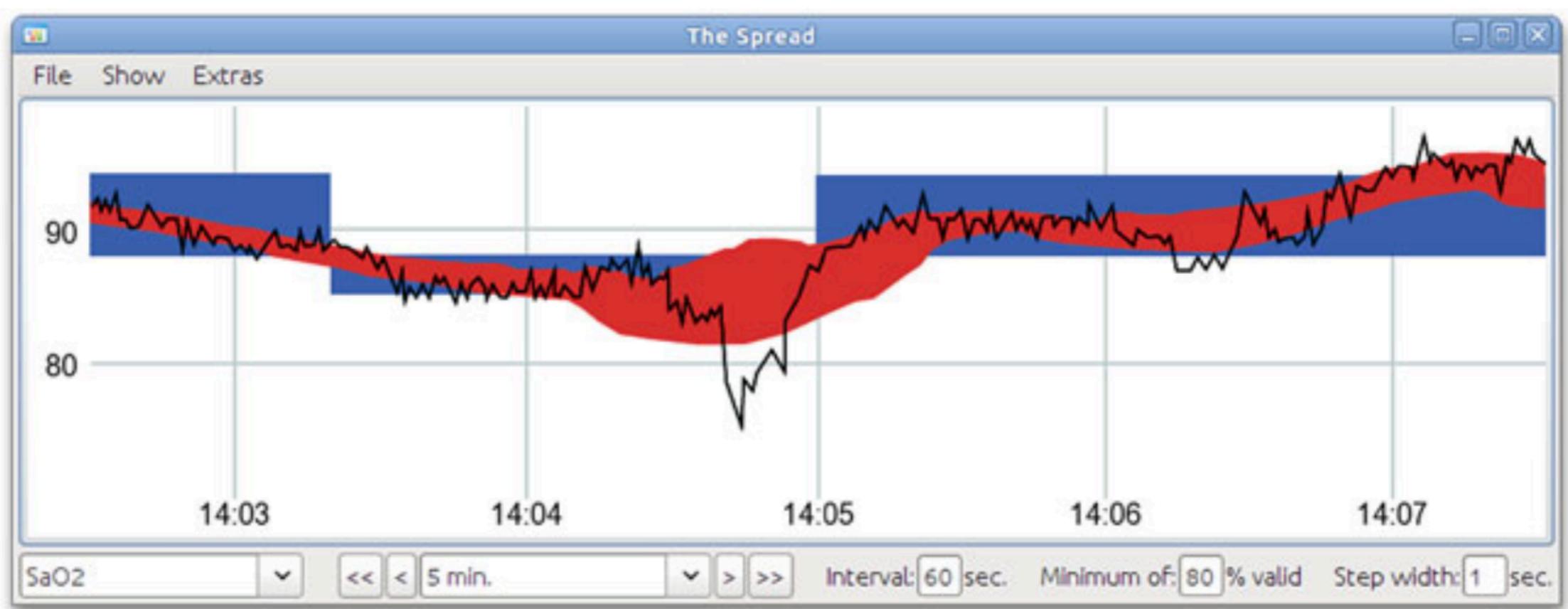
Trendalyzer/Gap Minder



from: Aigner et al., Visualization of Time-oriented Data (2011)

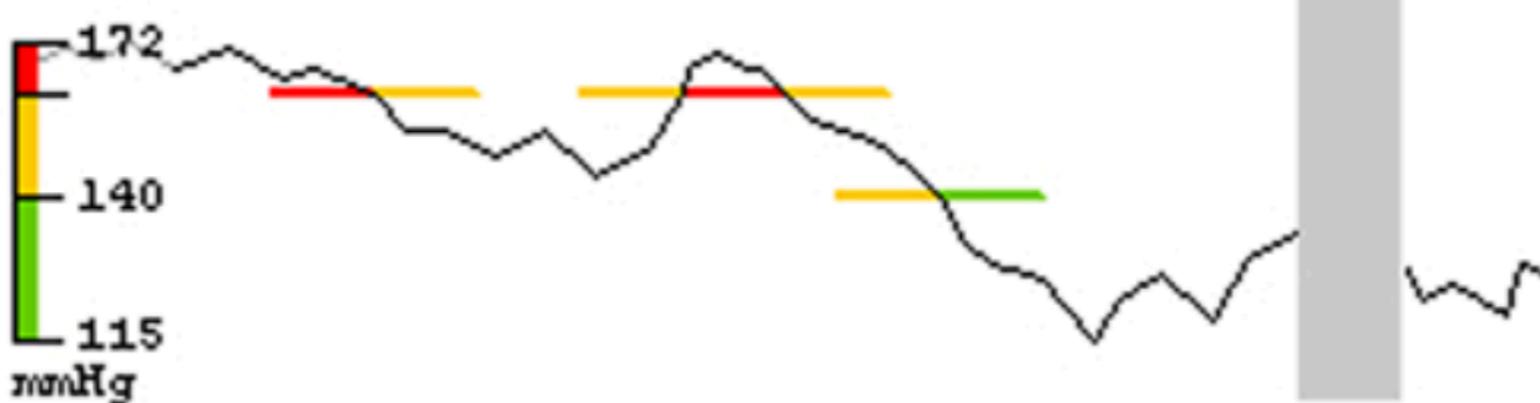
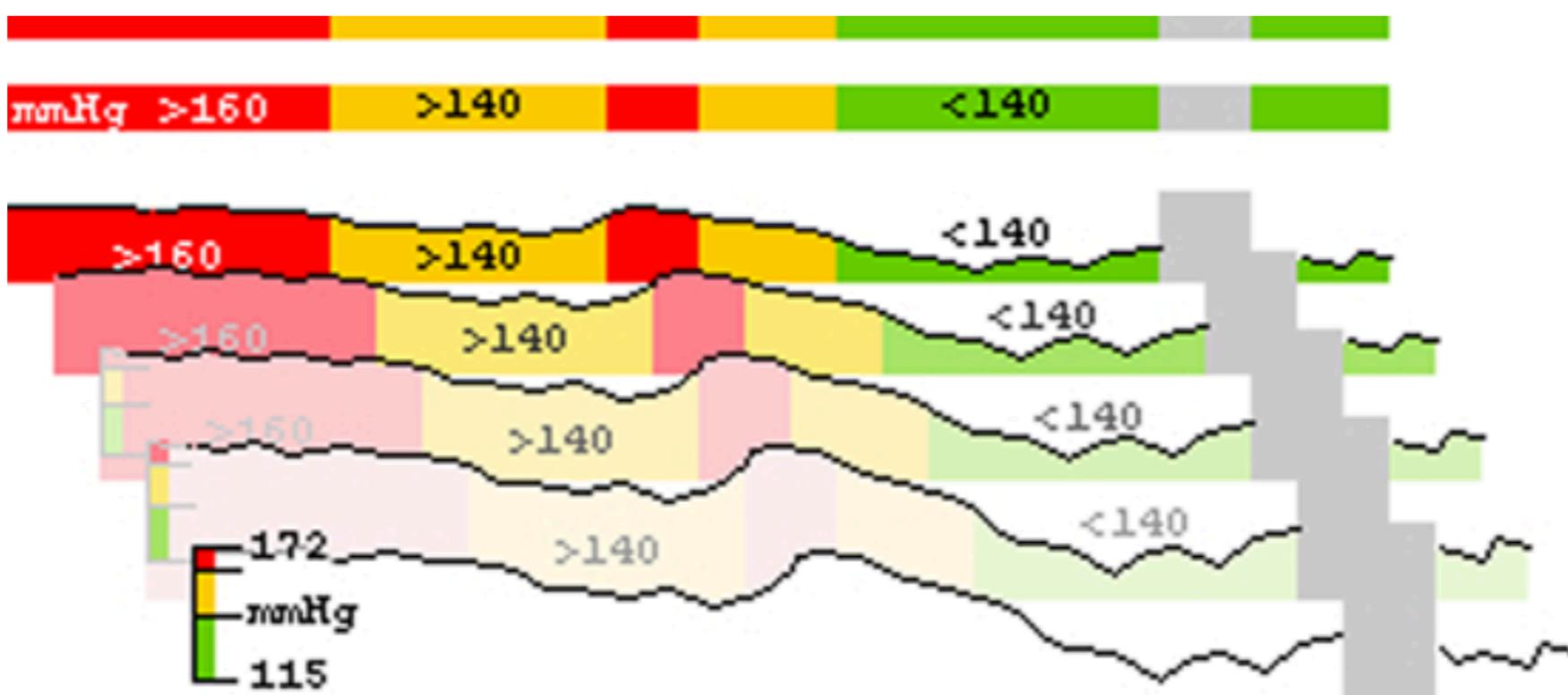
Smoothing and Scales

Dealing with Oscillating Data



Semantic Zoom with Oscillating Data

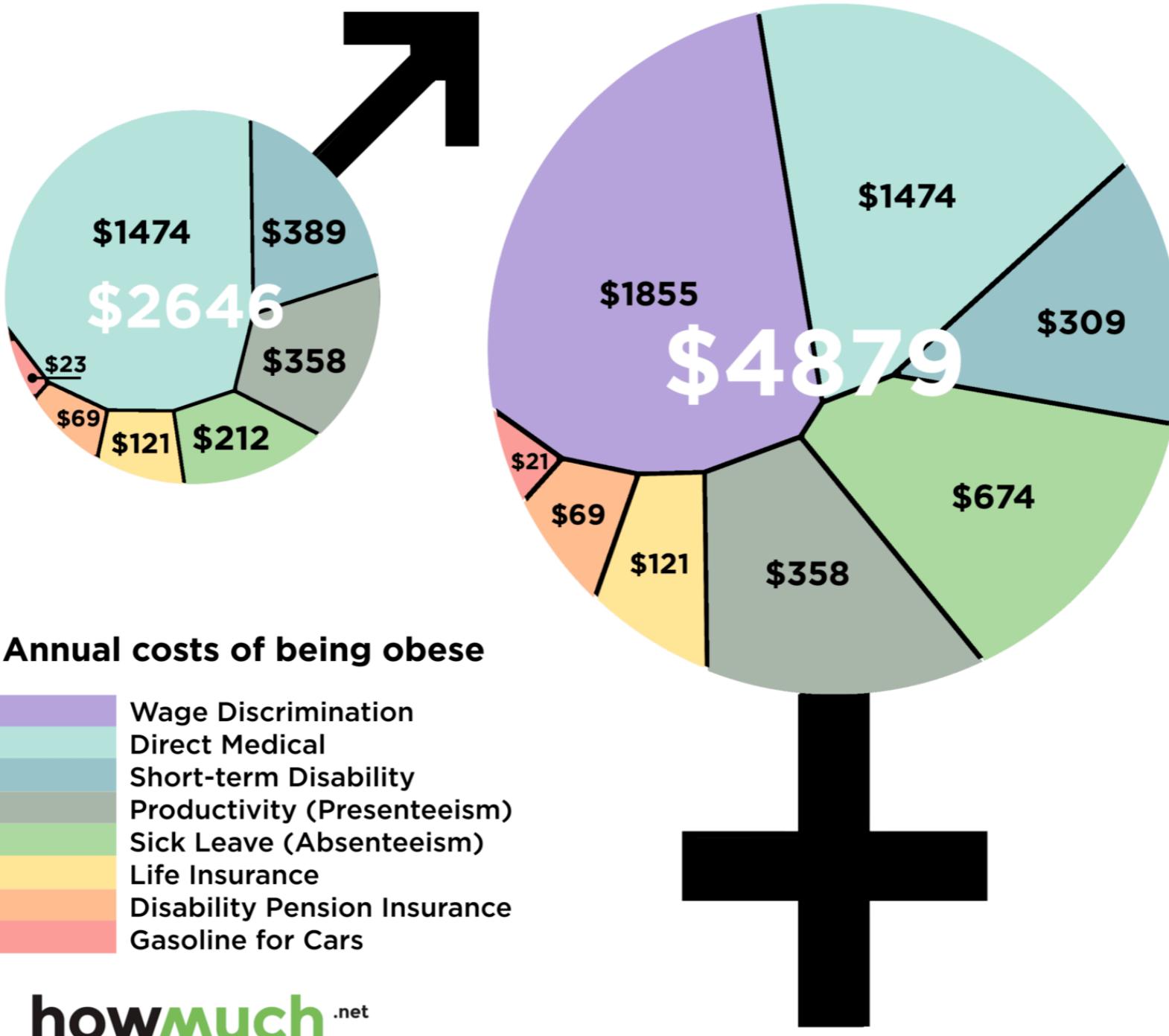
Zoom



Visualization for Communication

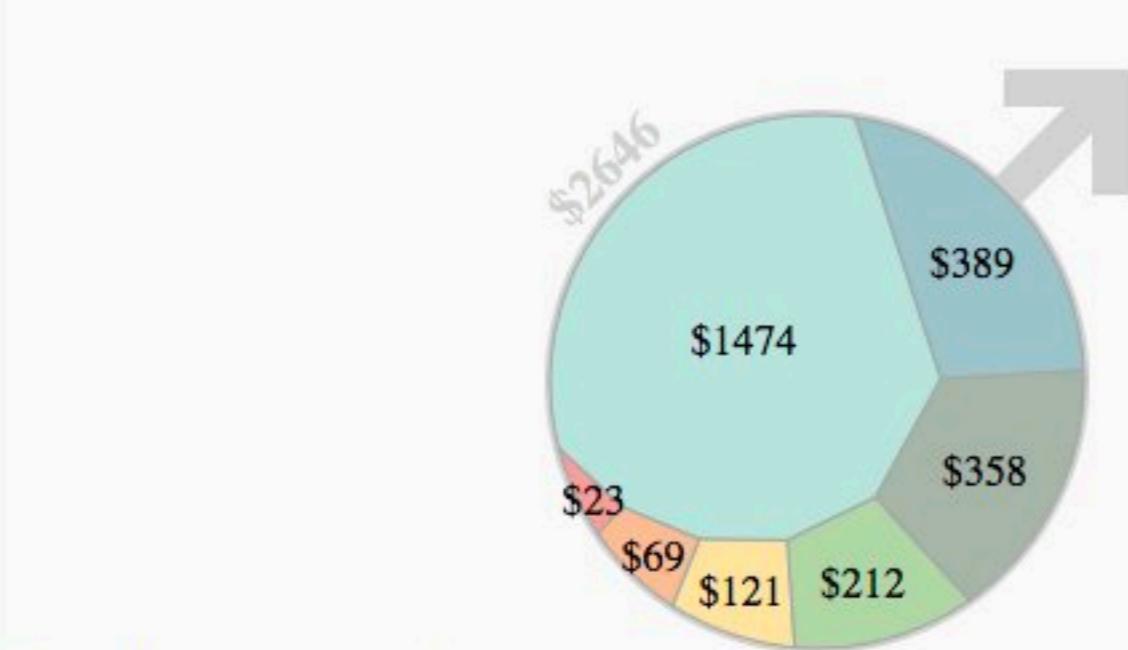
Redesign Exercise

The Individual Costs of Being Obese in the United States



Redesign Exercise

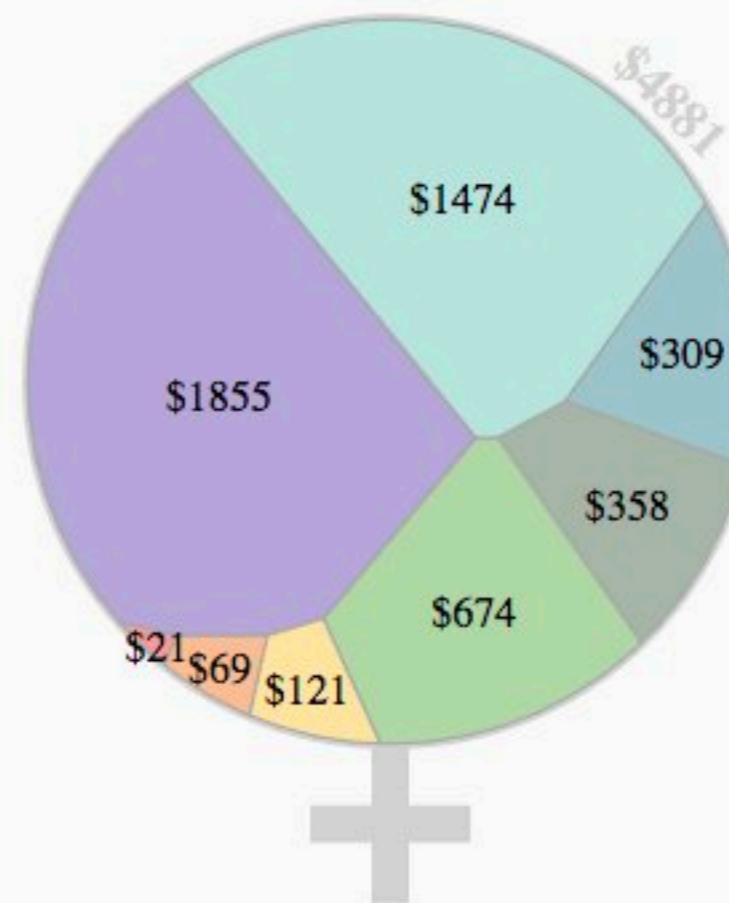
The Individual Costs of Being Obese in the U.S. (2010)



Annual costs of being obese

- Wage Discrimination
- Direct Medical
- Short-term Disability
- Productivity (Presenteeism)
- Sick Leave (Absenteeism)
- Life Insurance
- Disability Pension Insurance
- Gasoline for cars

Remake of HowMuch.net's post 'The Costs of Being Fat, in Actual Dollars'



by @_Kcnarf

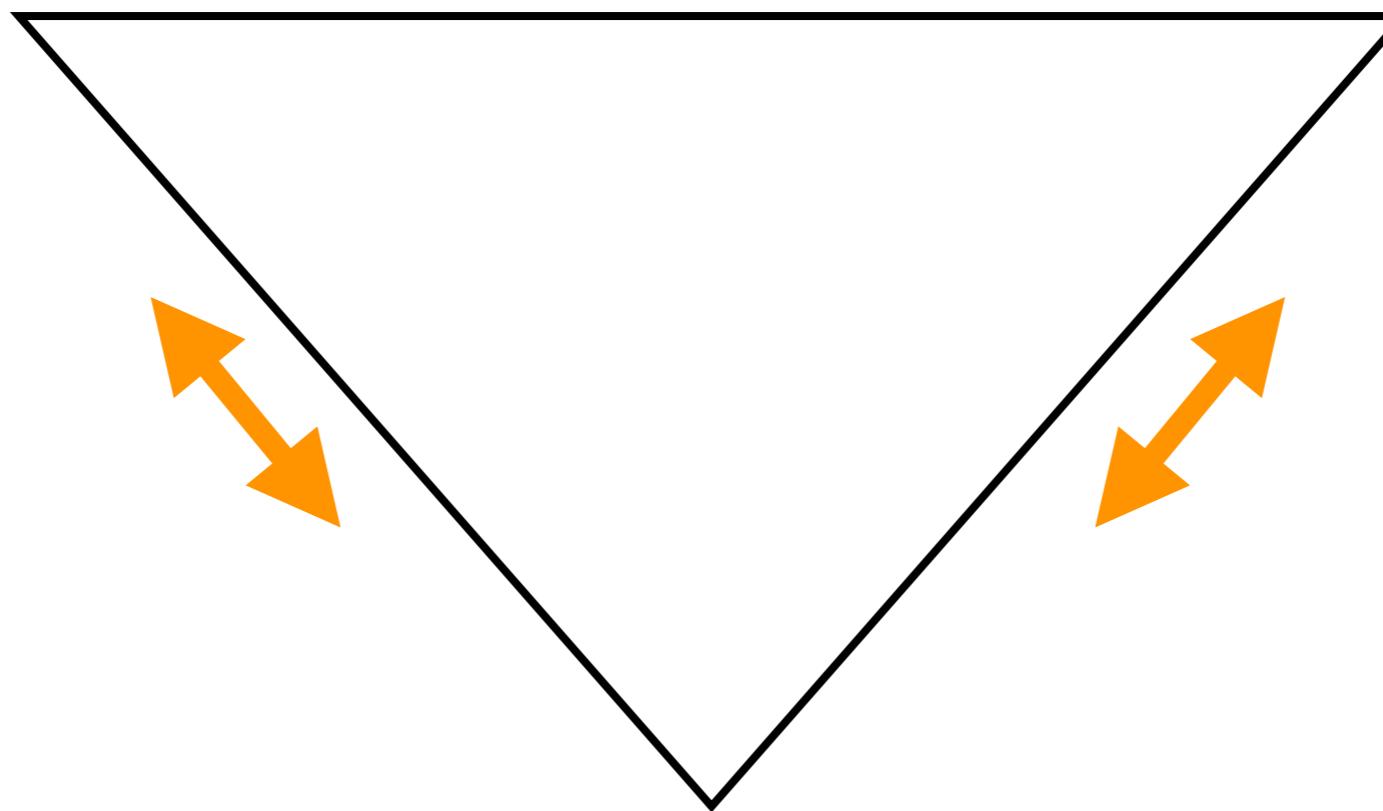
bl.ocks.org/Kcnarf/e649c8723eff3fd64a23f75901910930

dialogue between computer & analyst

Confirmation



Exploration



Presentation

dialogue between analyst & audience

Insights into the Paper, Data into the Database

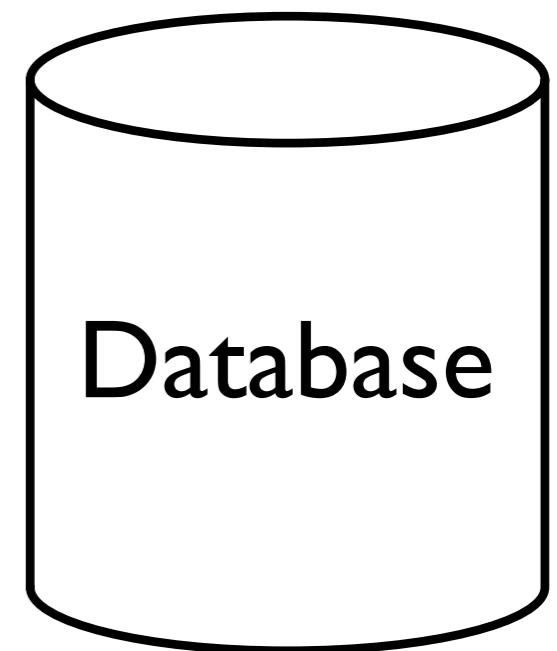
ARTICLE

doi:10.1038/nature11412

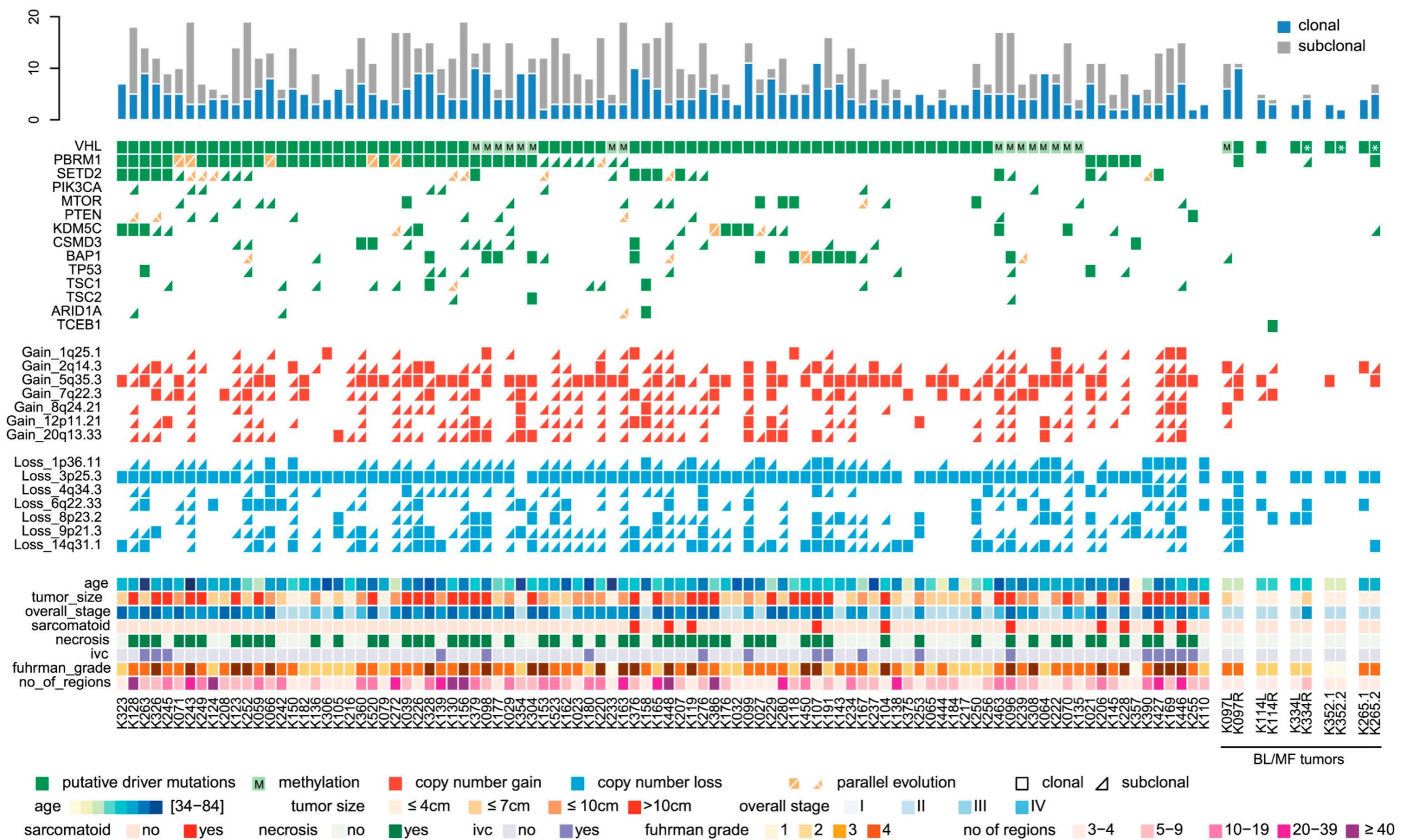
Comprehensive molecular portraits of human breast tumours

The Cancer Genome Atlas Network*

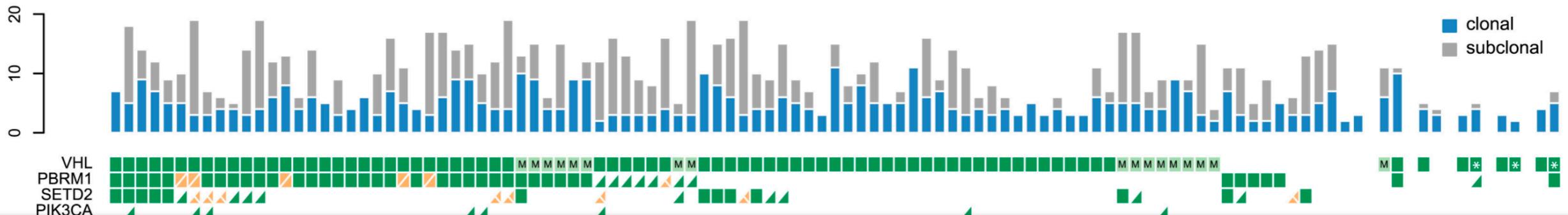
We analysed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. Our ability to integrate information across platforms provided key insights into previously defined gene expression subtypes and demonstrated the existence of four main breast cancer classes when combining data from five platforms, each of which shows significant molecular heterogeneity. Somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) occurred at >10% incidence across all breast cancers; however, there were numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in *GATA3*, *PIK3CA* and *MAP3K1* with the luminal A subtype. We identified two novel protein-expression-defined subgroups, possibly produced by stromal/microenvironmental elements, and integrated analyses identified specific signalling pathways dominant in each molecular subtype including a HER2/phosphorylated HER2/EGFR/phosphorylated EGFR signature within the HER2-enriched expression subtype. Comparison of basal-like breast tumours with high-grade serous ovarian tumours showed many molecular commonalities, indicating a related aetiology and similar therapeutic opportunities. The biological finding of the four main breast cancer subtypes caused by different subsets of genetic and epigenetic abnormalities raises the hypothesis that much of the clinically observable plasticity and heterogeneity occurs within, and not across, these major biological subtypes of breast cancer.



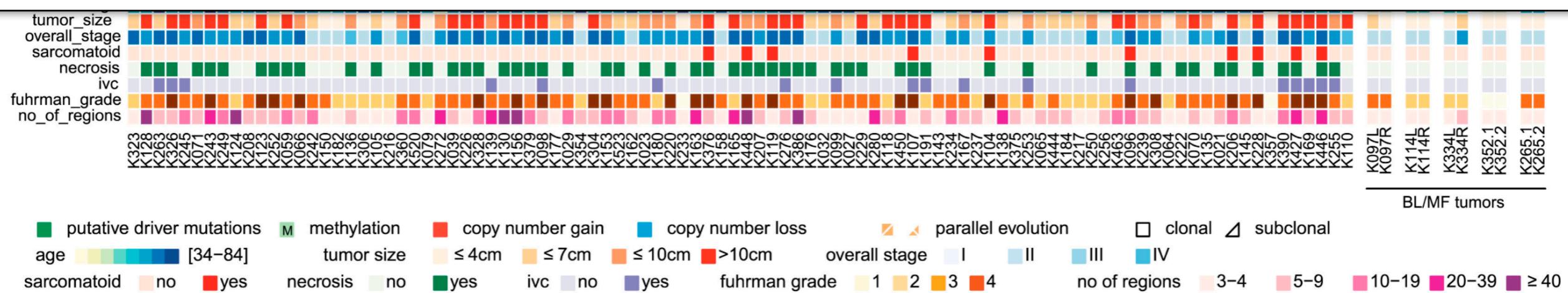
Insights into the Paper, Data into the Database



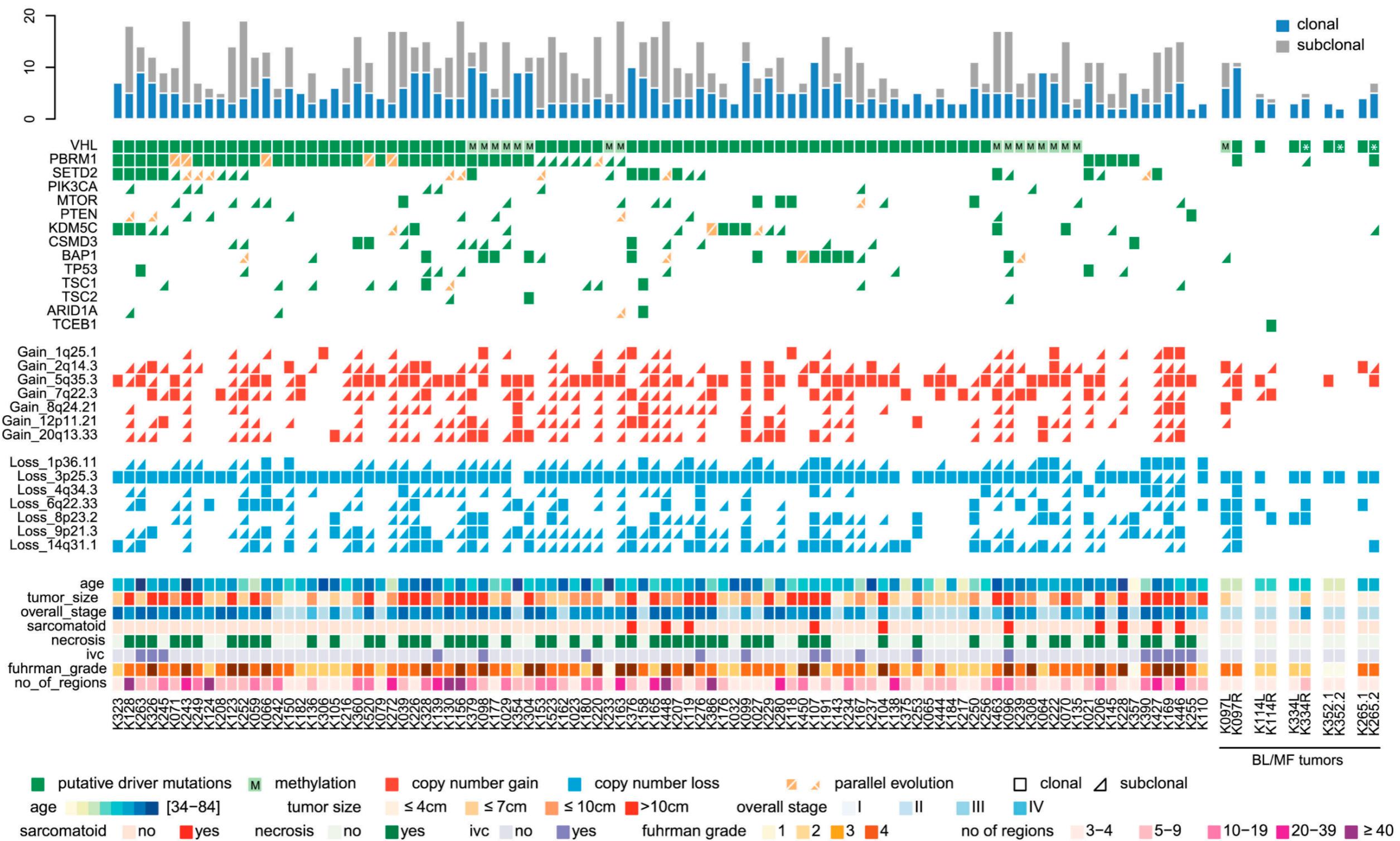
Insights into the Paper, Data into the Database



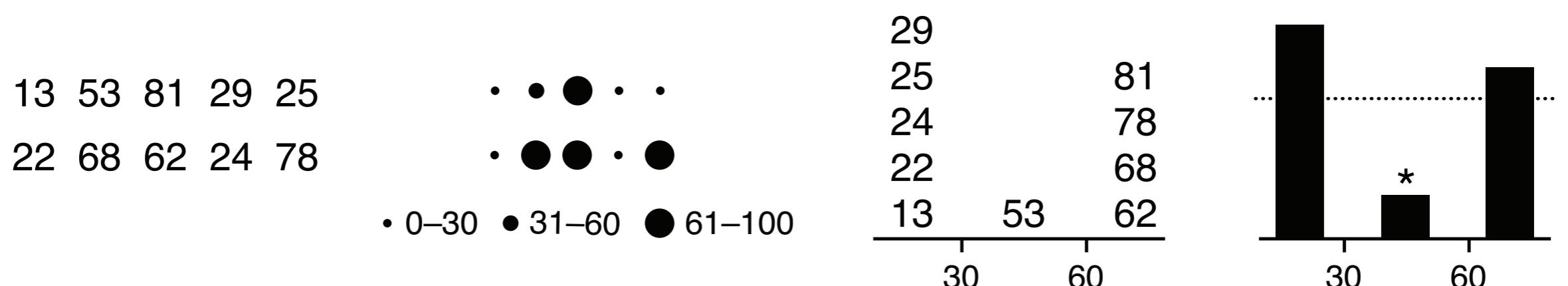
Overview of somatic driver alterations, including SNVs, DNVs, INDELS, and SCNAs, detected in the tumors of 101 TRACERx Renal cases. Rectangles and triangles indicate clonal and subclonal alterations, respectively. Parallel evolution is indicated in orange with a split indicating 2 or more parallel events. Five bilateral/multi-focal cases are shown on the right; distinct VHL mutations within tumor pairs are indicated with an asterisk.



Insights into the Paper, Data into the Database

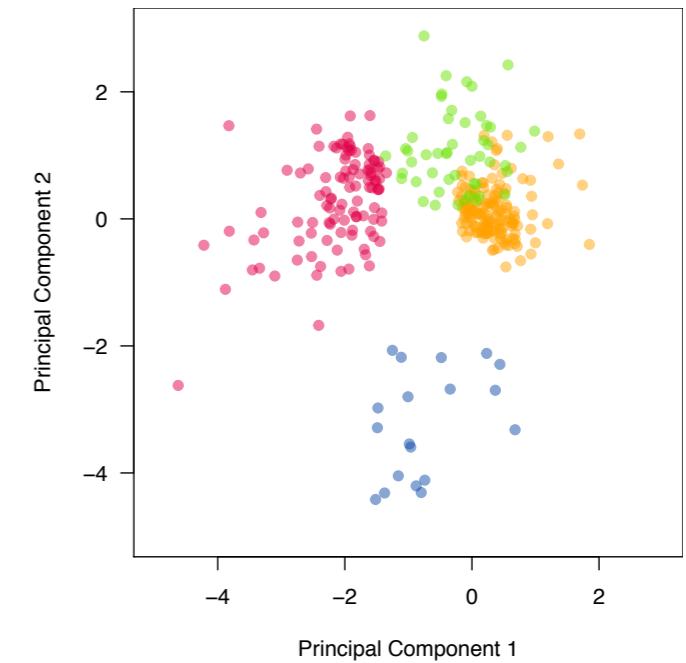
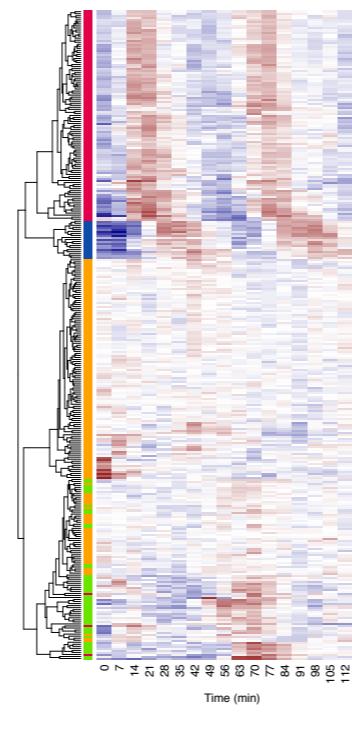
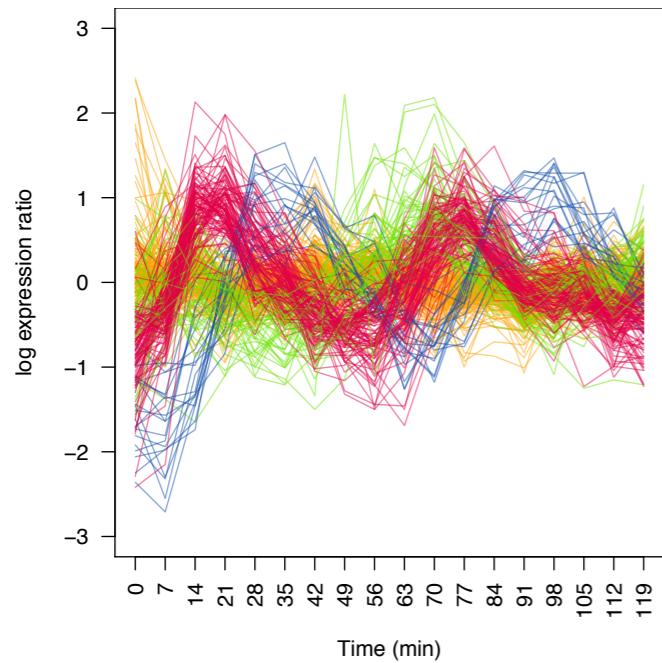


Focus on the Message: Reduce Data Detail



Focus

Choose a Plot Appropriate for your Message

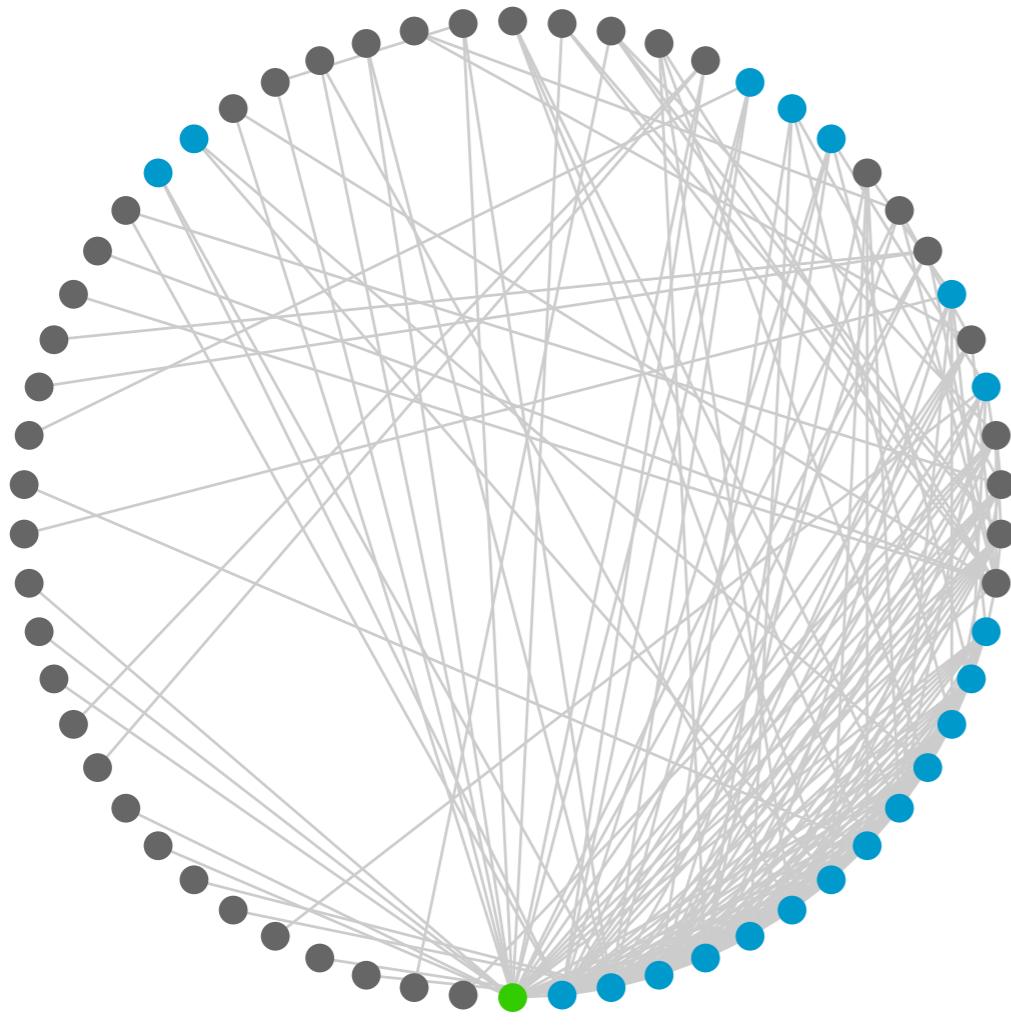


few, high-res

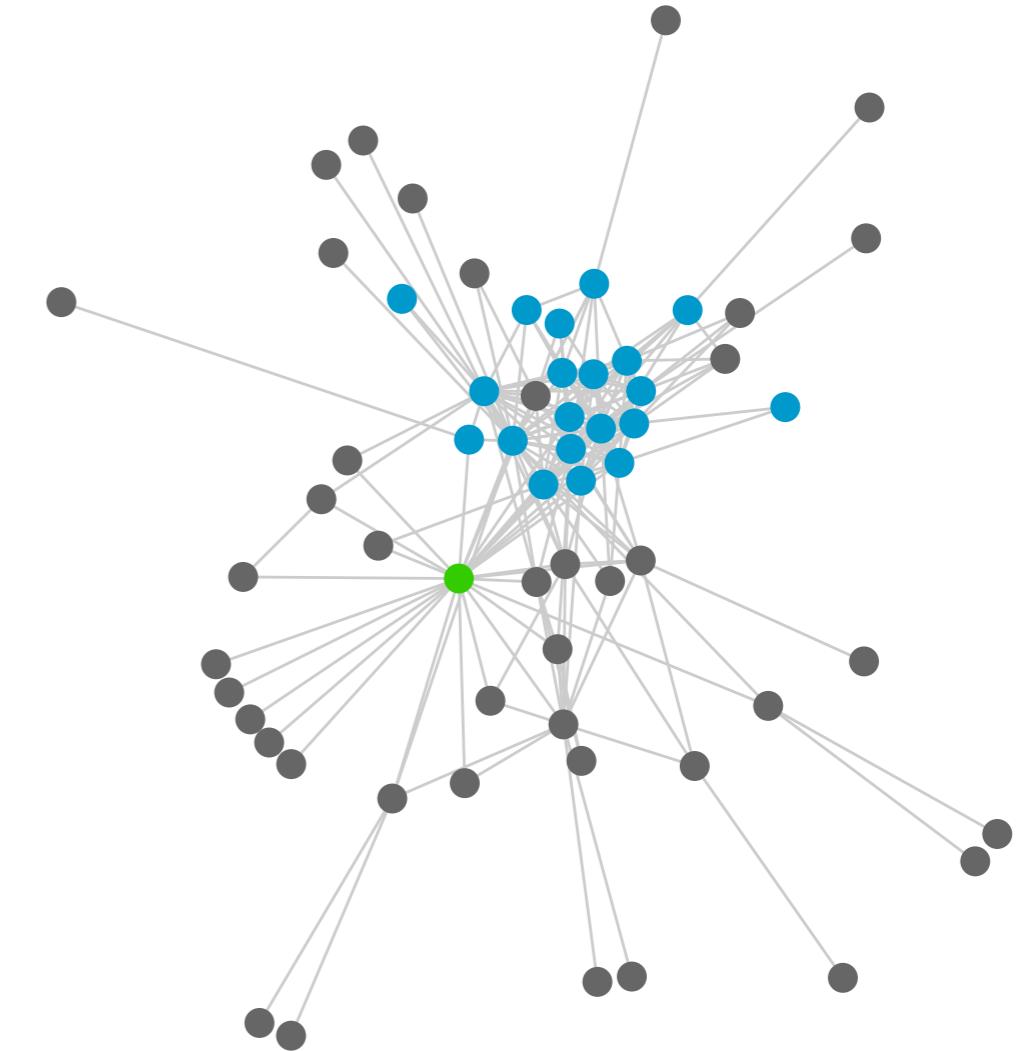


many, low-res

Choose a Plot Appropriate for your Message

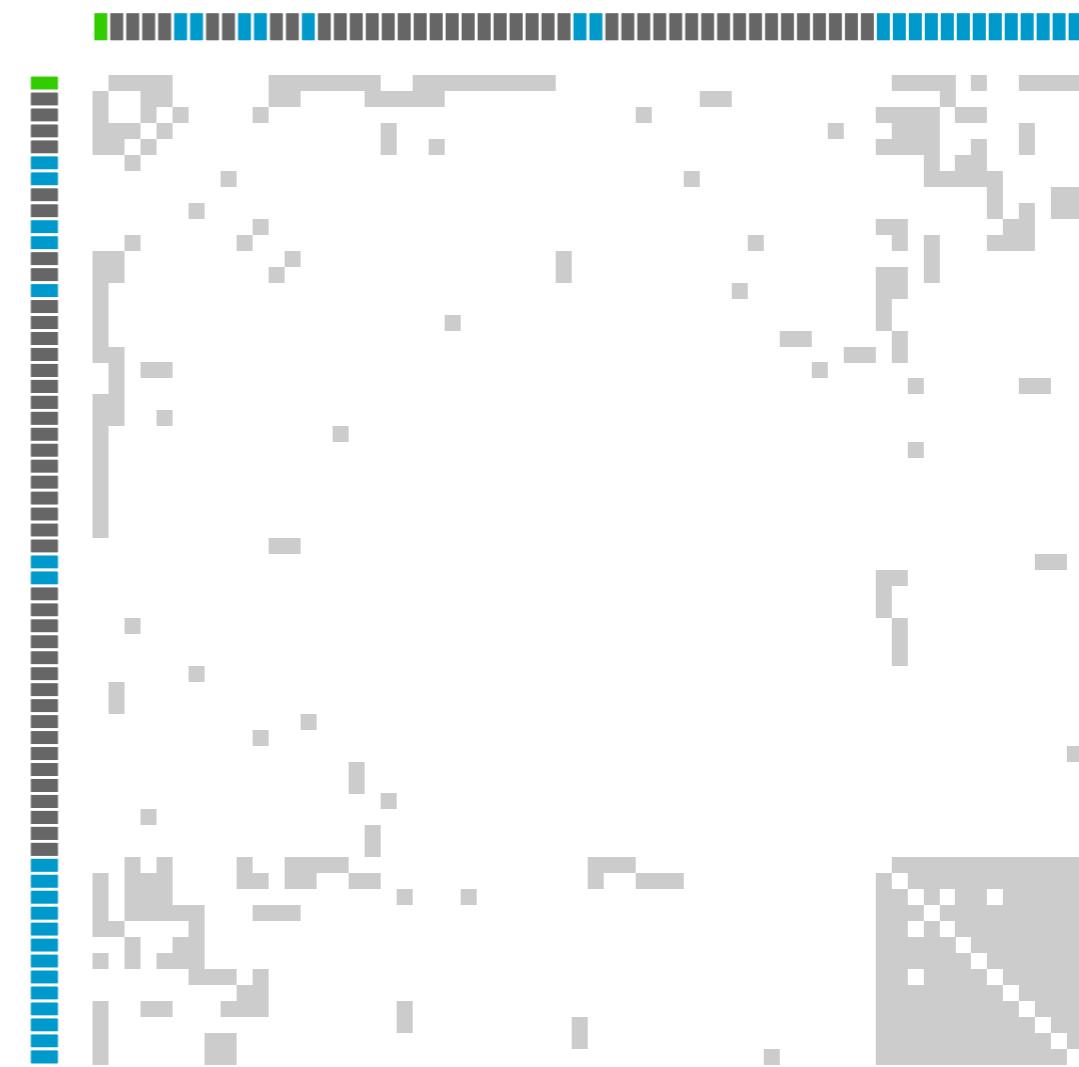
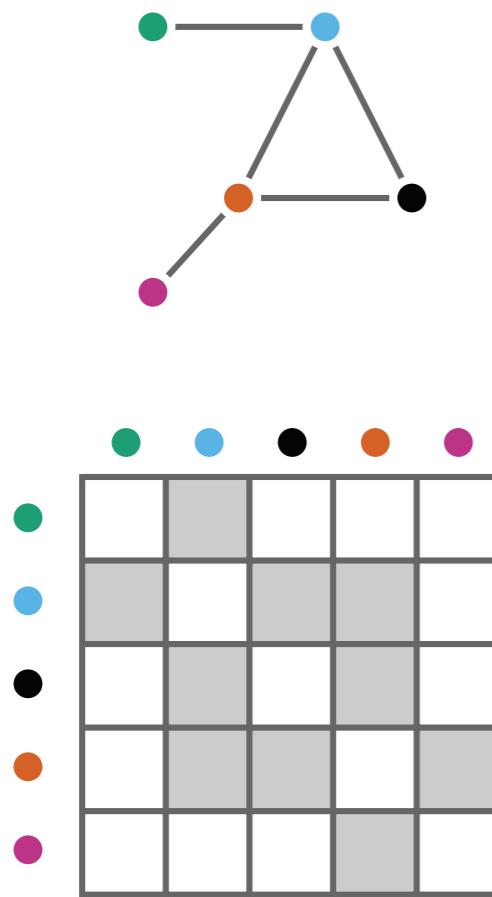


Circular Layout

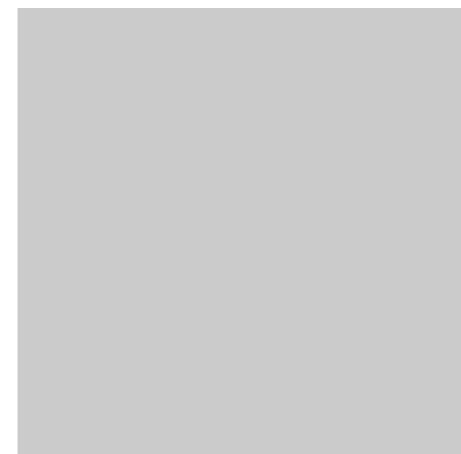
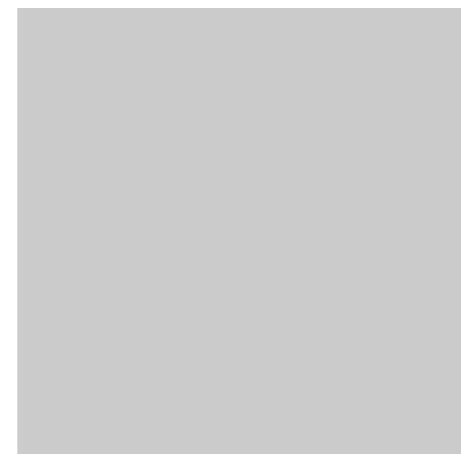
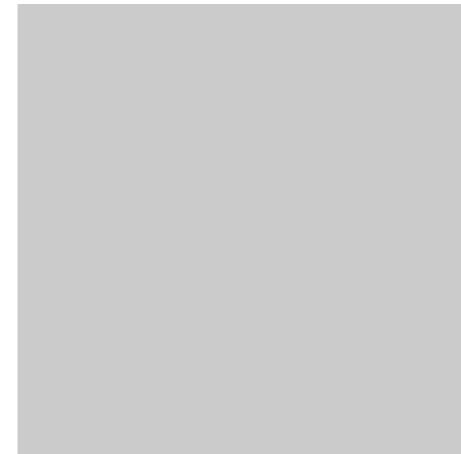
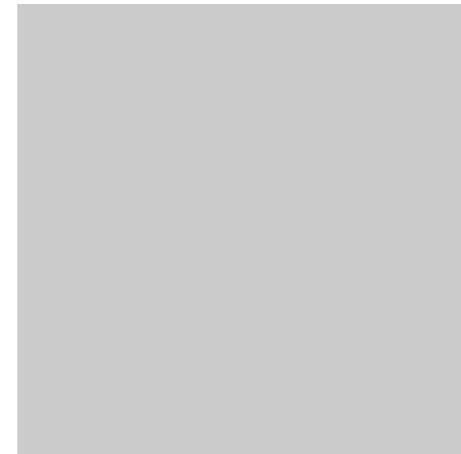


Force-directed Layout

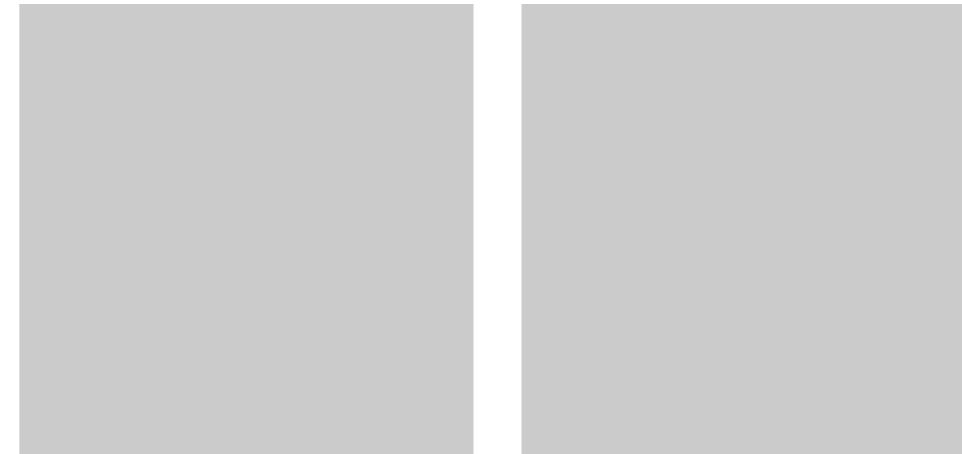
Choose a Plot Appropriate for your Message



Use Layout to Convey Meaning



Use Layout to Convey Meaning

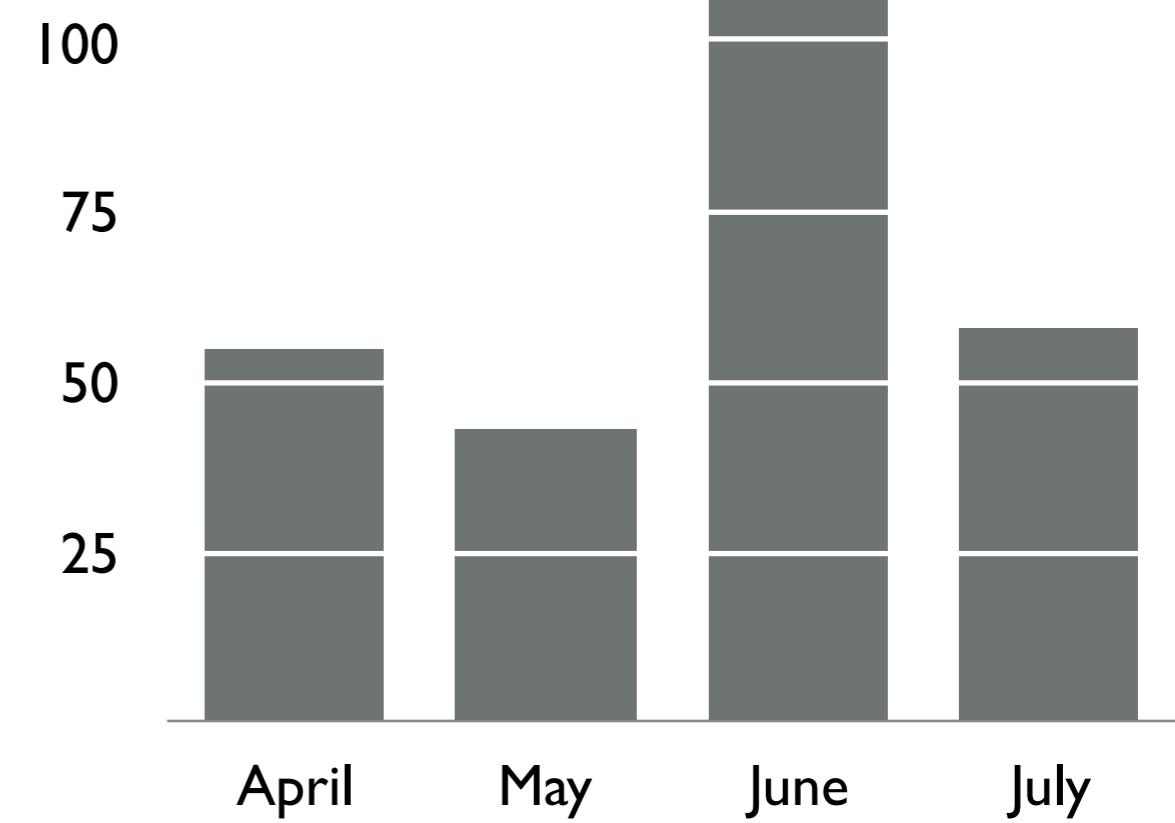
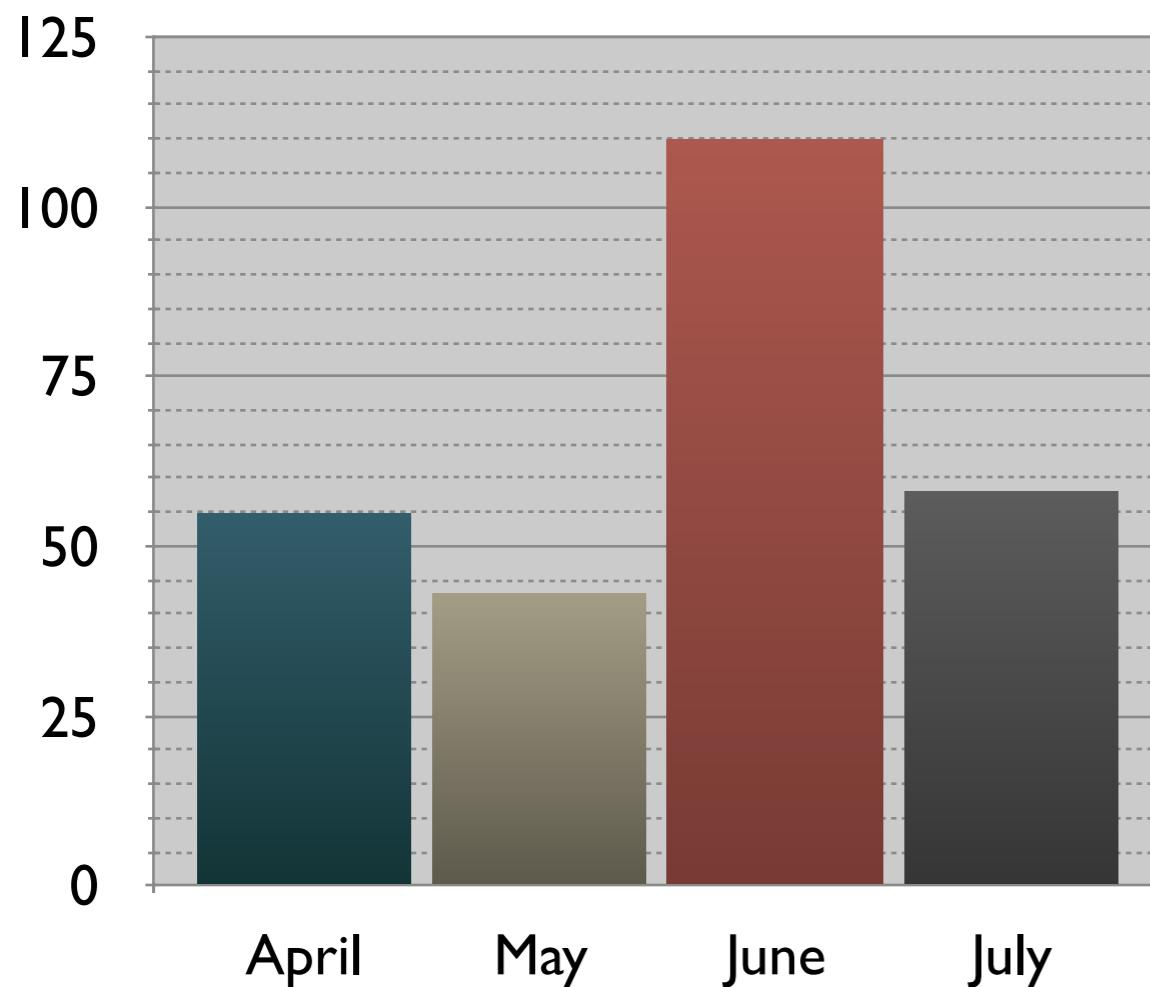


Use Layout to Convey Meaning

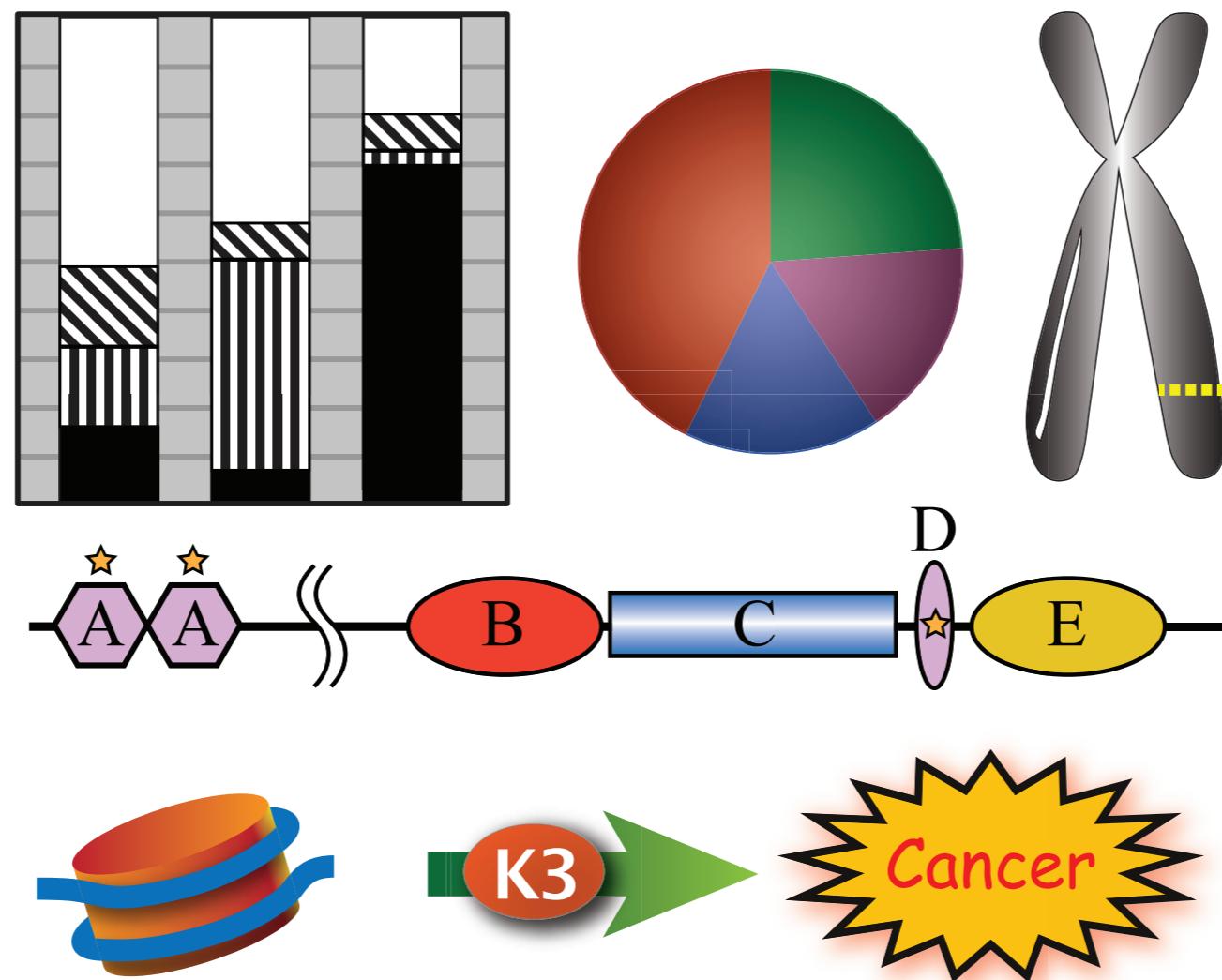


after Wong, Nature Methods, 2011

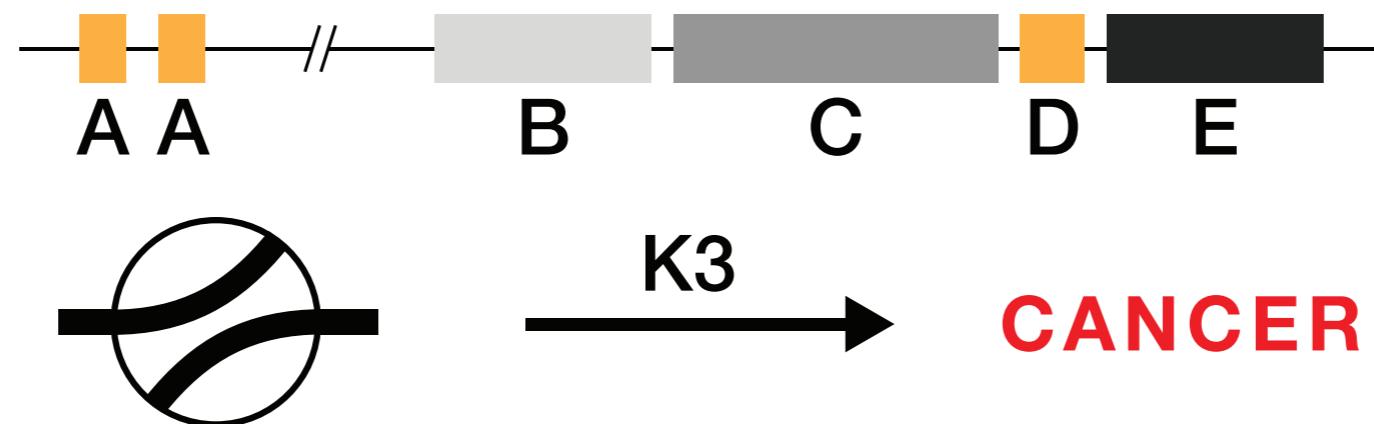
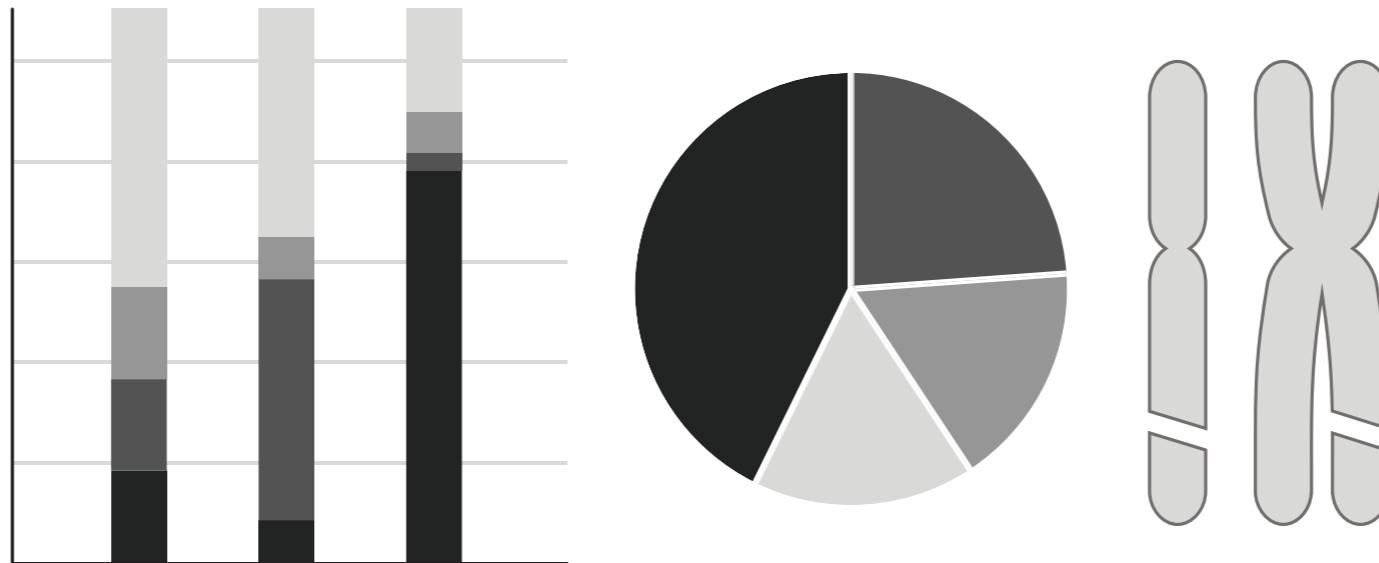
Focus on the Data: Maximize Data/Ink Ratio



Don't shout at your audience!



Don't shout at your audience!

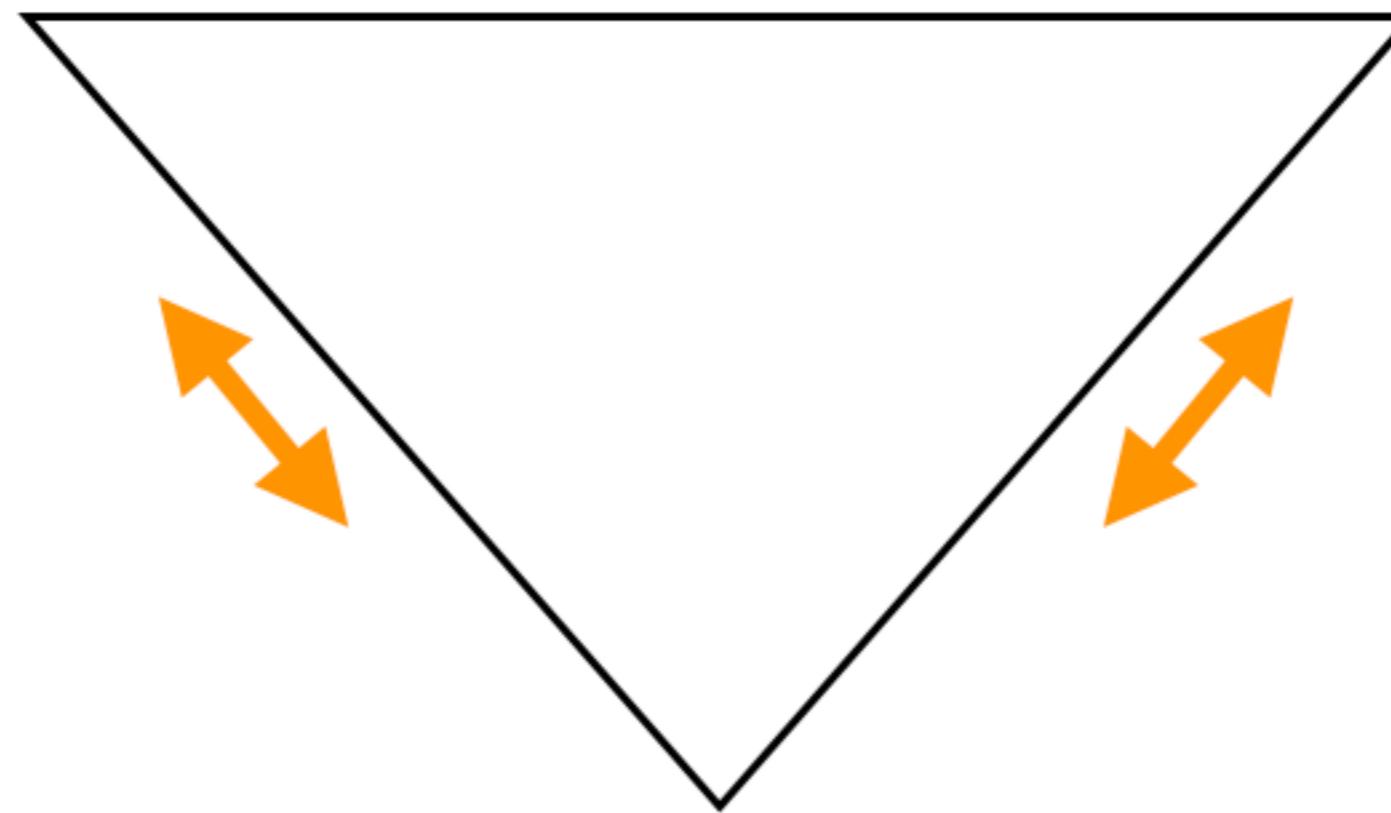


dialogue between computer & analyst

Confirmation



Exploration



Presentation

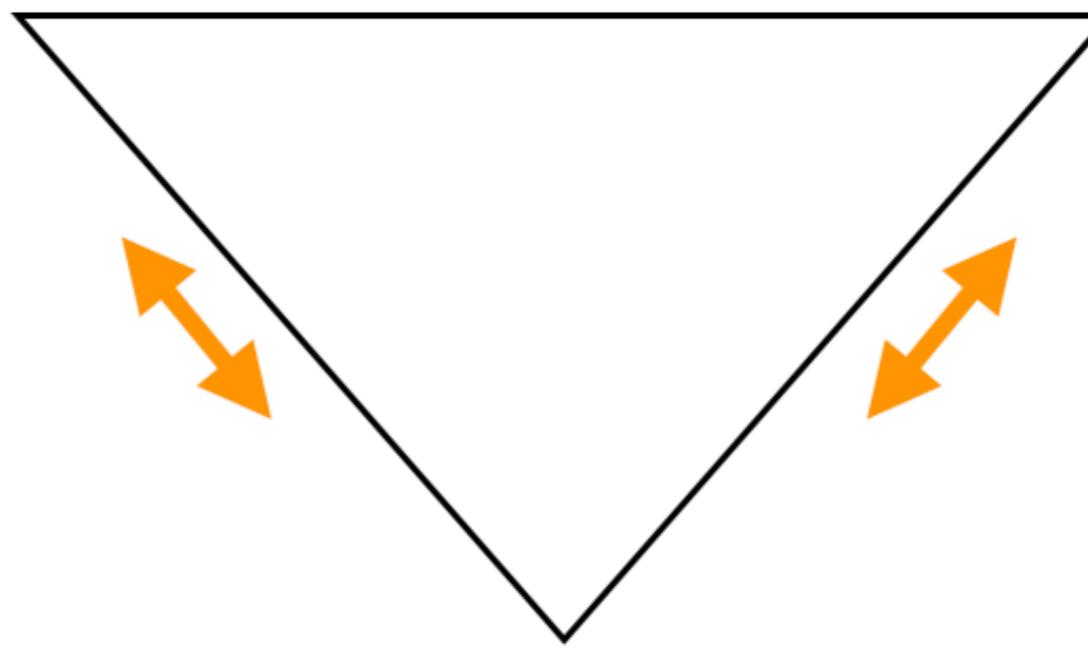
dialogue between analyst & audience

as much data as possible

dialogue between computer & analyst

Confirmation

Exploration



Presentation

dialogue between analyst & audience

as much data as necessary