Understanding and reacting to the impact of COVID-19 by combining statistics with visualization

Basel Biometrics Section Seminar: Graphics for decision-making in biomedical research and drug development

March 8, 2021

flatiron

Topics

- Flatiron Health and real-world evidence

- A principle we'll draw upon in using visualization to aid quantitative analysis: generate hypotheses

- Application: understanding the impact of COVID-19 on Flatiron Health data

flatiron

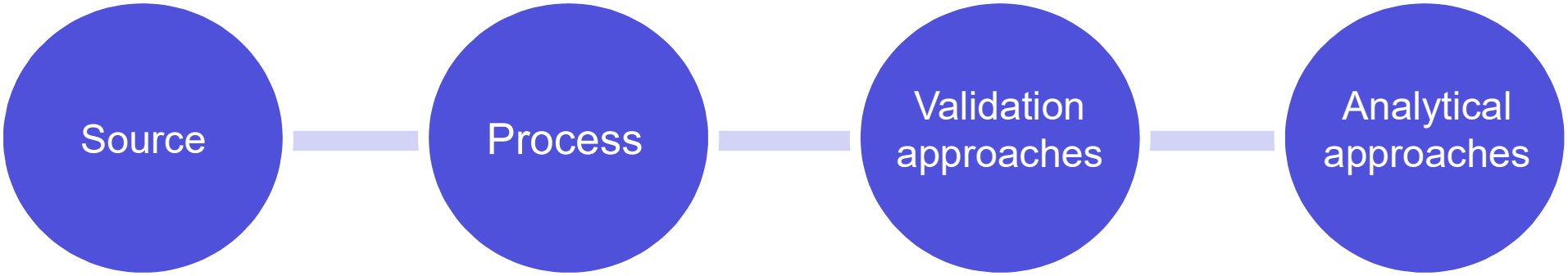
Real-world data and real-world evidence

- **RW data** are ***data*** relating to patients health status and delivery of healthcare routinely collected from various sources

- **Real-world evidence (RWE)** is derived from real-world data (RWD) through the application of research methods

flatiron

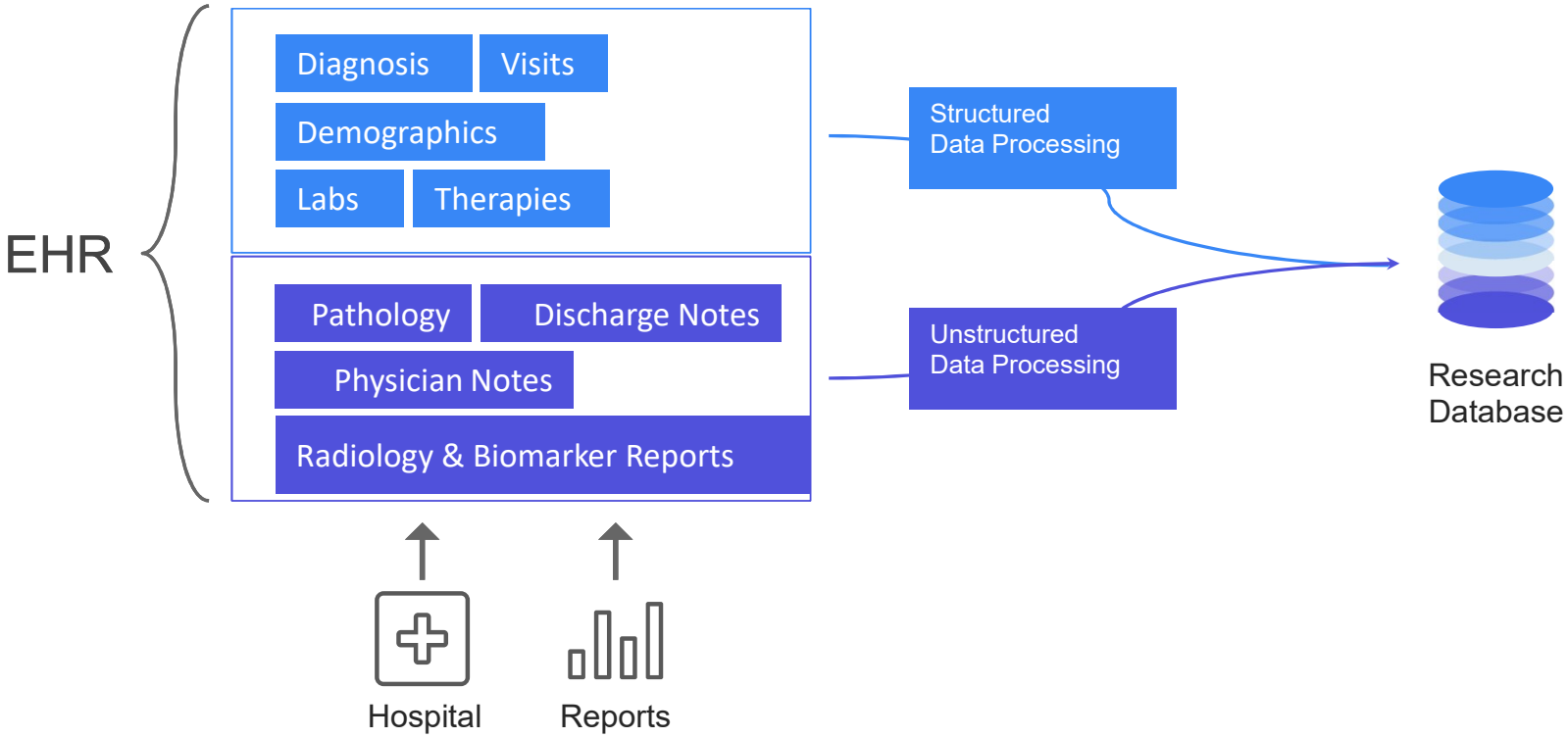
Flatiron's approach: developing high-quality real world evidence



"RWE is evidence derived from RWD through the application of research methods"

flatiron

RWD curation: structured and unstructured data



flatiron

Topics

- Flatiron Health and real-world evidence
- A principle we'll draw upon in using visualization to aid quantitative analysis: generate hypotheses
- Application: understanding the impact of COVID-19 on Flatiron Health data

flatiron

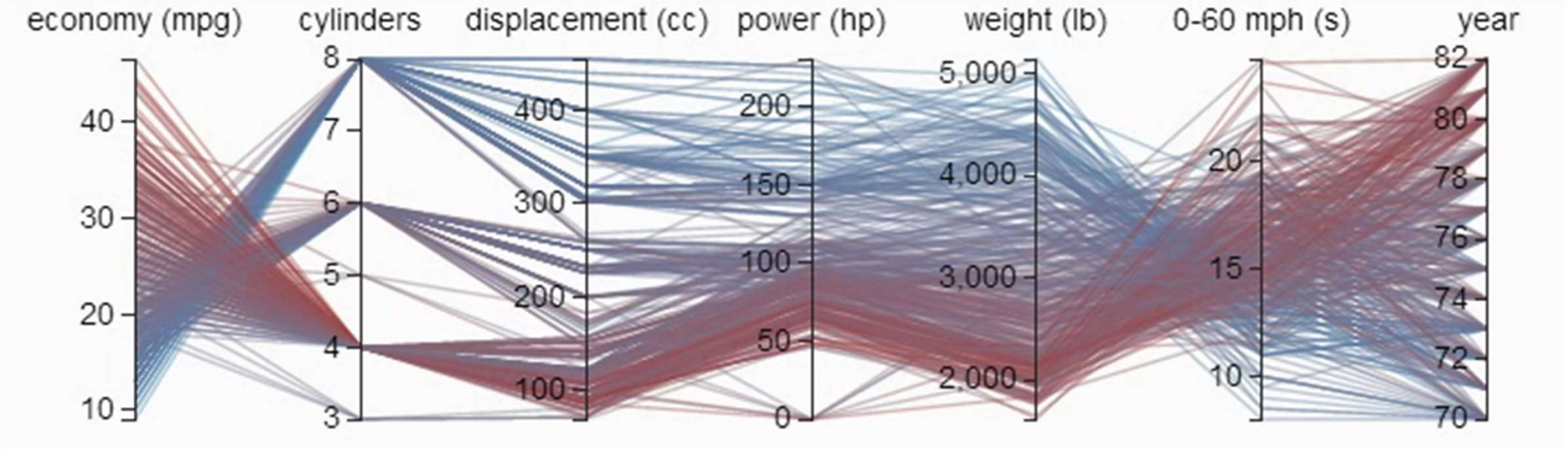
A question we'll hear: if we'd like to understand whether two variables are correlated… why use visualization? Can we not simply compute the correlation statistic?

- **Imagine a 'vehicle' database table* with the following variables:**
  - Economic performance (mpg)
  - Cylinders
  - Displacement
  - Power (hp)
  - Weight (lbs.)
  - 0-60 mph (s)
  - Year
- **An analytical question may be:** is the power (hp) of a vehicle correlated with its economic performance (mpg)?
- **Yet just one analysis here may lead to deeper questions:** which dimensions are correlated, and under what conditions?
  - We could prepare a correlation statistic on each possible pair of variables…
    - Yet… what if a given correlation only emerges among vehicles designed prior to 1990…
      - That weight less than 4000 lbs…?
        - **And what if we're not certain of the pertinent subsets before we begin?**
          - Strictly-computational approaches exist that will search for high-performing model specifications…
          - Yet… what if we have outside knowledge that should be incorporated into this choice of a specification?
          - Some computational approaches *will* also allow us to incorporate hypotheses and other tacit information (e.g., by beginning the search in the a-priori believed most-promising areas)
          - **Interactive visualization allows us to formulate and update these hypotheses in real time,** potentially avoiding unhelpful – and perhaps even infeasible – computation

Example on next slide

**in this example, we'll utilize the 'Motor Trend Car Road Tests' database (built into the R programming language)*
Reference: *Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.*

Example



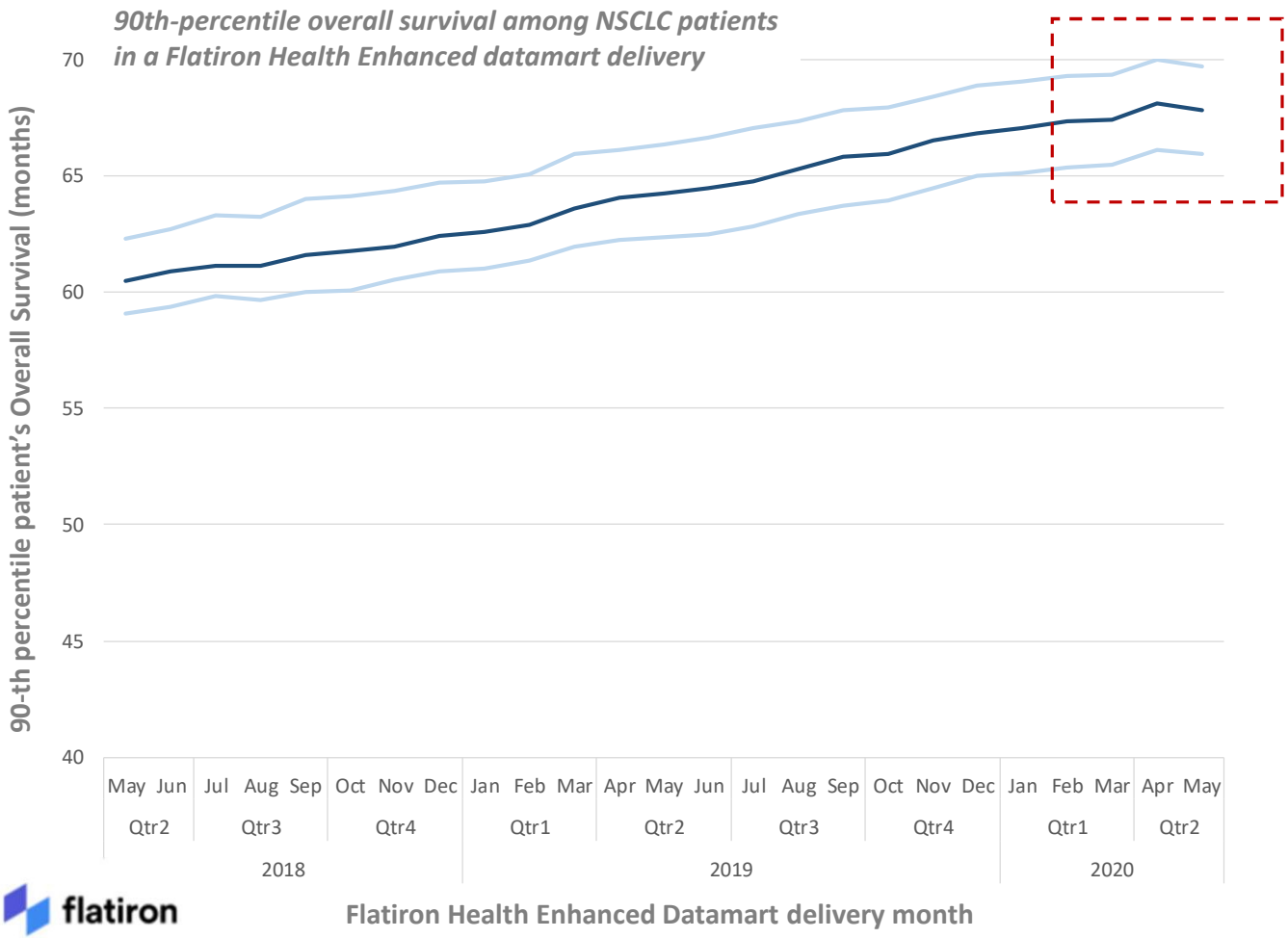
Reference: *Parallel Coordinates Toolkit,*
http://syntagmatic.github.io/parallel-coordinates/

Topics

- Flatiron Health and real-world evidence
- A principle we'll draw upon in using visualization to aid quantitative analysis: generate hypotheses
- Application: understanding the impact of COVID-19 on Flatiron Health data

flatiron

In April and May of 2020, we noticed that the survival of patients in our advanced NSCLC datamart – under certain ways of measuring this survival – began to break trend



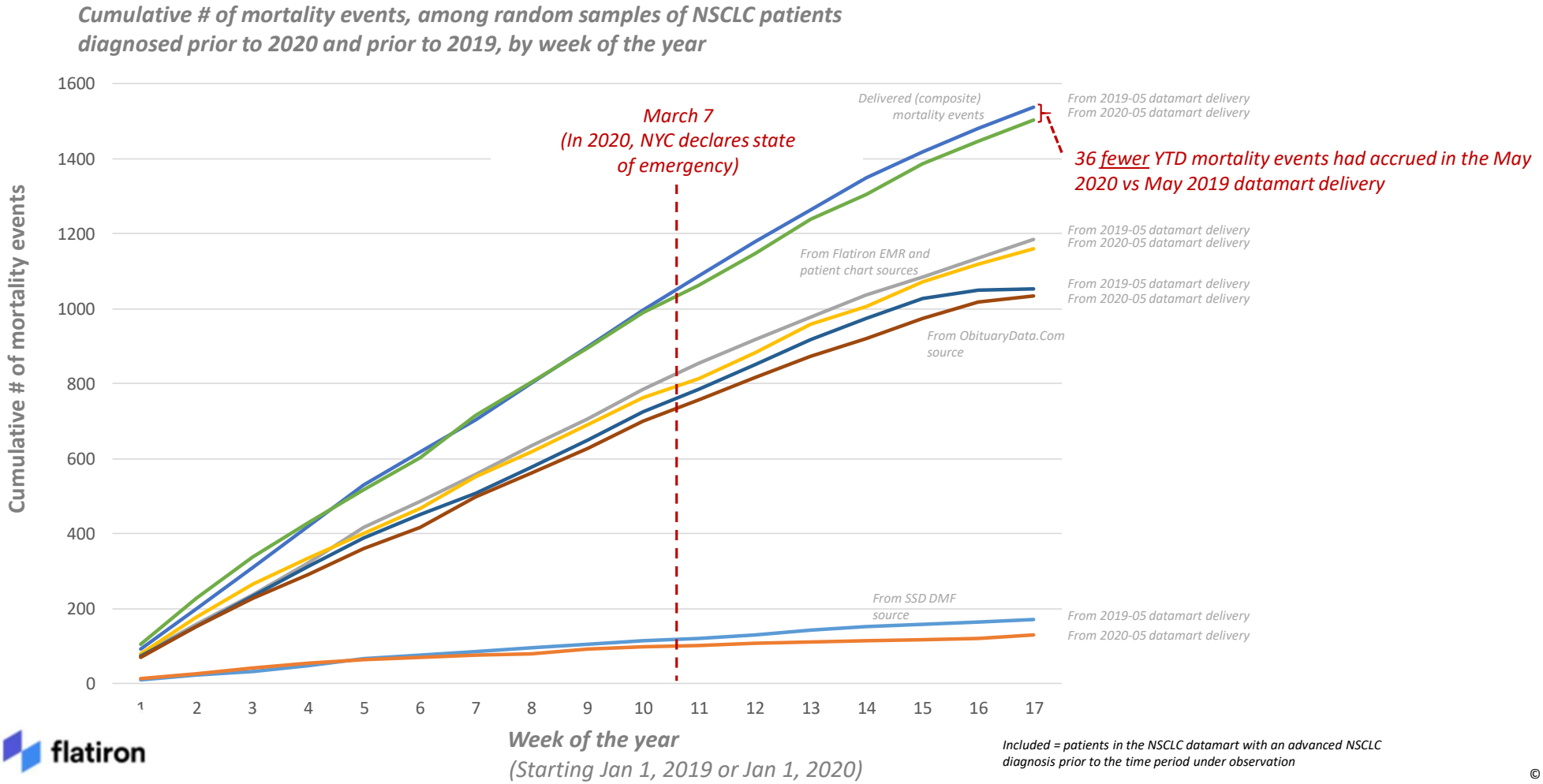
Is NSCLC patients' survival affected by COVID-19?

What is happening?

flatiron

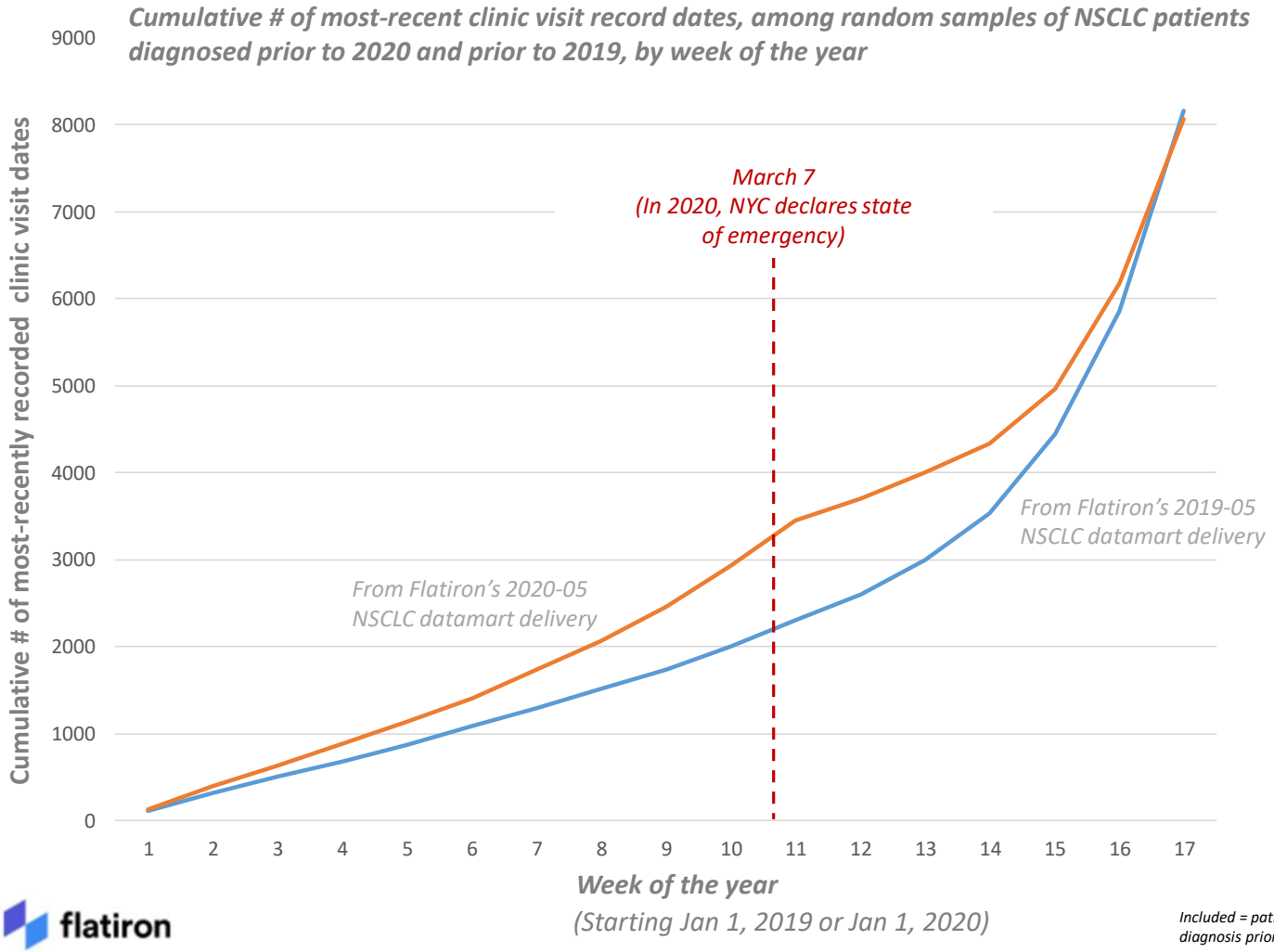
Hypothesis sequence	Visualization-driven finding	Decision
1. A COVID-driven increase in mortality was reducing overall survival among NSCLC patients	Slightly *fewer* mortality events per NSCLC patient were actually present in our data than over the same time period one year prior	Explore patterns in any additional variables that can contribute to overall survival calculations
2. Something unusual is happening with the data that one may use for longitudinal censoring	Patients' most-recently recorded clinic visit dates were in fact skewing earlier	

Our first hypothesis was that a COVID-driven increase in mortality was reducing overall survival. **A visualization of the accrual of mortality events among samples of patients from our May 2020 versus May 2019 datamarts showed that this was not actually as imaginable**



Included = patients in the NSCLC datamart with an advanced NSCLC diagnosis prior to the time period under observation

Second, we looked at the distribution of patients' *most-recently recorded clinic visit dates* which, under certain approaches to calculating overall survival, play a role in longitudinal censoring



The May 2020 datamart's distribution of patients' most-recently recorded clinic visit dates was skewed toward times prior to NYC declaring its state of emergency

Certain approaches to calculating overall survival will utilize a composite of a patient's most-recently recorded clinic visit date and date of death (if one exists) to censor patients from that calculation

Under such an approach, a slowing in the accrual of final visit records could have led to a small, but noticeable – and yet artificial – decrease in calculated overall survival

flatiron

Included = patients in the NSCLC datamart with an advanced NSCLC diagnosis prior to the time period under observation

In conclusion

- When calculated using Flatiron's May 2020 datamart, certain survival-related statistics were showing a **small but noticeable downward break from the prior trend**
- Exploratory visualization helped us see that our **first hypothesis** – that, overall, patients at our clinics were passing away sooner, conceivably as a result of COVID-19 – was **less imaginable than first believed**
- Instead, we saw that **changes in patient visit patterns were more likely to be driving these trends** in measured statistics
- From here, we were able to make certain that any **further statistical analysis was performed with this dynamic in mind**

flatiron