Process Book

Basic Info.

Project Title:

Emergency Medical Complaints

Team:

Ada McFarlane

email: u1087069@utah.edu

uID: u1087069

Patrice Nicholes

email: Patrice.nicholes@m.cc.utah.edu

uID: u0073711

Repository:

GitHub: https://github.com/lateoclock/dataviscourse-pr-EmergencyMedicalComplaints

Background and Motivation

As biomedical informatics students, we are interested in using data to improve healthcare. Patrice is a Research Analyst at the Utah Department of Health, Bureau of EMS and Preparedness. She regularly works with pre-hospital data, as well as various health data repositories. BEMSP has a goal to return more of the data obtained from local agencies back to those agencies with added value in order to promote participation in data sharing programs. This is especially important for small, rural agencies and facilities that don't have in-house data analysis resources. This visualization could be a means of showing state program participants that their data is being used. It would be especially successful if the recipients find it fun and easy to use. Patrice will obtain EMS and emergency data from UDOH.

When a person calls in to 911, they are sometimes panicking or rushed. The caller may not be the person in distress, or the person in distress may be unable to respond to questions from dispatchers. Although dispatch follows a protocol to best document the callers complaint, the complaint often does not accurately and completely capture the real problem. For example, a person may call in complaining that they are having breathing problems. The emergency medical team may prepare to treat a person with asthma, or a blocked airway, only to arrive at the scene and find a patient with anxiety having a panic attack. Urgency is defining characteristic of pre-hospital care. EMTs, AEMTs and paramedics need to assess and care for a patient as quickly as possible. Awareness of unexpected but likely possible diagnoses may help a care provider better anticipate needed treatment. Analysing outcome data from the hospital emergency department where the patient was eventually treated can add insight that EMS agencies could use in performance improvement and education efforts.

Project Objectives

The objective of this project is to produce a data visualization that will provide insight into the triage and treatment of patients that require emergency care. This visualization will answer questions like “when a person calls in to 911 complaining of a headache, what are possible outcome diagnoses based on previous patient calls?” and “among all of the patients diagnosed with a substance abuse disorder, what were the complaints provided to dispatch? What were the ratios of each complaint?” The visualization will also give a layperson a snapshot of how medical issues are documented across the continuum of care. For example, it will show how 44 dispatch complaints result in thousands of diagnoses in the emergency room. The user will learn whether a complaint generally leads to a single, straightforward diagnosis, or if it’s a symptom with complex etiology. The web based format is a benefit because it is more accessible to volunteer agencies that may not have facilities, hardware, or software. The visualization will give agencies and hospitals a chance to see the data they provide in a new way and allow them to compare what they have noticed in their own organization and what is happening across the state as a whole.

Data

The data for this project will come from pre-hospital and emergency department data from 2017. More recent data is not yet available. The UDOH BEMSP has recently started to link pre-hospital data with data from emergency departments in hospitals across the state. The data will not contain any PHI. This will somewhat limit possible filtering and sorting options, but including PHI would complicate access tremendously. There will be three fields: dispatch complaint, EMS provider primary impression, and emergency department discharge diagnosis. There is one row per provider impression/diagnosis combination per incident. The dataset contains approximately 40,000 rows, 44 categories for complaints, 419 types of provider impressions, and 3500 diagnosis codes. Combinatorial explosion will be an issue and may be handled with sorting, filtering, consolidating, and data clean-up. We may also choose complaints or diagnoses that are of particular interest and only include those data points.

Data Processing

Because the data in 2017 was not validated to the extent that is now, it will need to be cleaned up substantially. There are some missing values, typos, and lack of terminology

standardization. There is also some messiness introduced with matching the emergency department dataset with the pre-hospital data set, mostly in the form of not matching a patient incident in both sets. This can be mitigated somewhat with careful data queries. The data will need to be formatted so that it can be easily rendered into our diagrams and charts. Our program will need to do some aggregation. The main quantities to be derived from the data are counts and ratios of complaints for given diagnoses. Data processing will be implemented with a combination of SQL queries, some spreadsheet maneuvering, and javascript.
The size and complexity of the data led to

The two views in our visualization required different formats for the data in order to work:
For the Sankey diagram: We needed to have the data arranged as a set of nodes with unique ids, and links between those nodes with each link being between 2 nodes rather than a 3-node path. To get this format for the data, we used the pandas library for Python in Jupyter Notebooks to process the data into a manageable size and then output the resulting subset in the required format. We made several attempts at this, and eventually were able to get a viable dataset for the model by omitting all paths that occurred fewer than 10 times and then choosing the top 5 most common complaints, the top 5 most common impressions for each of those complaints, and the top 5 most common diagnoses for each of those impressions.
For the bar chart: This view is a little more flexible when it comes to data formatting, and we ended up using d3's nest and rollup methods to aggregate the raw list of incidents (our starting format) into a nested dictionary of complaints, impressions associated with those complaints, and the frequency counts for each of those impressions. From that point, we were able to sort the impressions for each of our complaints of interest by frequency and generate a bar chart showing the dropoff rate. Originally, we had intended to implement a Pareto line on the chart to emphasize this feature, but we didn't get that feature finished in time and had to scrap it.

Visualization Design

The data could be displayed in a tree and corresponding tree map, as a network diagram, a chord diagram, or a heat map. Some possible solutions include:

- The root node of the circular tree could be a 911 call, with the first level of nodes being the dispatch complaint, the second being the provider impression and the third being ED diagnosis. The tree could expand and collapse nodes. The terminal node could be selectable to provide a filter criterion for another visualization.
- A tree map that can show the ratios of dispatch complaints represented for any given diagnosis or impression selected with a search bar.
- A chord diagram could trace the path from dispatch to EMS to discharge. The different complaints could be highlighted on hover.
- A heat map in the form of a human body could be used to display the frequency of illness and injury to a particular part of the body, with dropdown menus that let the user choose the level of care.

Our ideal visualization will be a network/tree diagram and a heatmap represented as a patient's body. This will depend on the feasibility of obtaining data on body regions.

Must-Have Features

This visualization must represent the three types of data, complaint, provider impression, and ED diagnosis and the relationship they have with each other. It should have interactive filtering or sorting in order to make the large amount of data manageable.

Optional Features

Including appealing design, color and illustration in addition to the actual data would be a bonus. For example, putting the tree map into an image of an ambulance as the back of the truck may add interest and rememberability. Highlighting a path on the tree diagram from start to finish would make it easier to distinguish from the rest of the tree. An indicator of a selected value on a tree node, such as larger text, bold color, or change of the color of the node would help users see the selection better. Color saturation and hue could add emphasis or code for additional features.

Project Schedule

October 18th - Team Announcement. Ada sets up the repository and turns in the completed form.

October 25 - Ada and Patrice have met to brainstorm. Patrice submits proposal.

November 1 - Patrice acquires data and data is clean, formatted, and ready for use. Ada will lead the structural design of the code. Ada and Patrice have completed the display of the basic layout and form of the project. There is a date scheduled to meet with the TA for the review.

November 8 - Patrice and Ada have met with the TA and worked out any bugs. Tasks are divided up between Ada and Patrice. Milestone one is submitted. Begin designing and publishing website.

November 15 - Interactivity is added to the visualization. The project website is up and running with links to the repository. Begin work on the screen cast, including basic script. Edit and finish the process book.

November 22 - The screencast is filmed and added to the readme file in github.

November 27 - Peer review is completed and project is submitted.

<u>Data Vis Final Project Checklist</u>

- ❏ Data Formatting
    - ✓ Top5 (Sankey) - data subsetted & export into usable format
    - ✓ All Data - derive frequencies/counts & export into usable format:
        - paths-oneplus.csv = paths with freq <1
        - paths-twoplus.csv = paths with freq <2
    - ❏ Assign categories for diagnosis types
    - ❏ Convert csv data to nodes & links json format for Sankey

1. Infection:
    - Sepsis
    - Other Infection
2. Pain:
    - Chest Pain
    - Other Pain
3. Substance Abuse
4. Injury:
    - Fractures
    - Other Injury
5. Respiratory:
    - COPD
    - Other Respiratory
6. MCI (incl. STEMI & NSTEMI)
7. Mental Illness:
    - Depression
    - Suicidal Ideation
    - Other Mental Illness
8. Weakness

- ❏ Organize Github Repo into folders (OldFiles, Reference, JupyterNotebooks, etc.)
- ❏ Time permitting - redo top5 to not include the “not recorded” values

- ❏ Questions we’re trying to answer:
    - ➢ When a given complaint is sent to EMTs, what might they expect to find when they arrive at the patient? (are complaints a good indicator of impression)

    - ➢ When an EMT documents an impression, what is the final diagnosis likely to be? (are impressions a good indicator of diagnosis?)
    - ➢ When a complaint is given to dispatch, what are the likely final diagnoses? (are complaints a good indicator of diagnosis)

- ❏ Design & Layout
    - ❏ Views :
        - ❏ Sankey -> Overview of flow
        - ❏ Bar Chart(s) with Pareto line -> show spread of data, especially to give context to Top5
        - ❏ Contingency table(s)? Alternate view shows relationship between 2 steps rather than all 3
    - ❏ Data Shown :
        - ❏ Top5
        - ❏ Time permitting - top5 without “not recorded”
        - ❏ Query (by keyword or by category?)
    - ❏ Controls :
        - ❏ Keyword Search Bar (do you think people will know what to search for?)
        - ❏ Query By category?
        - ❏ Controls for Bar Charts
        - ❏ Toggles?
        - ❏ Tabs??
    - ❏ Notation :
        - ❏ Project Title & Authors’ names
        - ❏ Attribution of sources
        - ❏ Label Charts & Diagrams
        - ❏ Explain what the views mean
        - ❏ Highlight interesting points (like how a lot of things are not recorded)

- ❏ Implementation
    - ● Sankey code from Jen: https://bl.ocks.org/GerardoFurtado/ff2096ed1aa29bb74fa151a39e9c1387
    - ❏ Sankey
    - ❏ Bar Charts w/ Pareto
    - ❏ “Verify” the html file : https://validator.w3.org/#validate_by_upload

- ❏ Hosting/Publishing
    - ❏ Get a domain
    - ❏ Publish & the site

- ❏ Documentation & Presentation
    - ❏ Codebook
    - ❏ Github Readme

❑ Video??

Feedback

Feedback from Jen:

“It is great you started setting up a visualization and it looks like you already processed the data. For the milestone, you need to have started the process book that you will be submitting for the final submission. It should have the background, motivation and questions you are asking about the data that are driving the visualization as well as the sketches you have made (before and after the feedback). It is helpful as well to include the feedback and the design choices you have made in response to the feedback. Looking forward to see the sankey diagram!”
JEN ROGERS , Nov 12 at 8:16pm

Feedback from class:

This project feedback was given by Fangfei Lan and Michael Young. They thought that EMS complaints are interesting to the target audience of EMTs and the general public, but acknowledged that the EMS audience may be interested in different aspects of the visualization than the viewers in class. The scope of the project is considered appropriate, but several suggestions were given that would add complexity to the design. The reviewers noted that they liked the circular design of one graph and that they have not seen many radial tree graphs. Fangfei suggested maybe using a sunburst graph instead. This would allow more use of color and allow for encoding of ratios. Graying out and collapsing were also recommended. Michael and Fenfei liked the tree map idea and suggested taking it a step further to zoom in on diagnosis rectangles in order to code for the primary impressions calegory, which otherwise is skipped over in the tree map representation. Scaling is a limitation of the data that is being used because there are many carepaths with the option of infinite carepaths. We discussed using regex to further categorize the data. Data could be limited to topics of most interest in the EMS community or to the most common care paths. They liked the story, and suggested using a story-telling element in the visualization, perhaps hover messages, or a box with statistics and other pertinent information. They also liked the idea of having an image of a human body with a heat map. We brainstormed ideas to add body regions to the data, including using regular expressions to categorise the diagnoses.

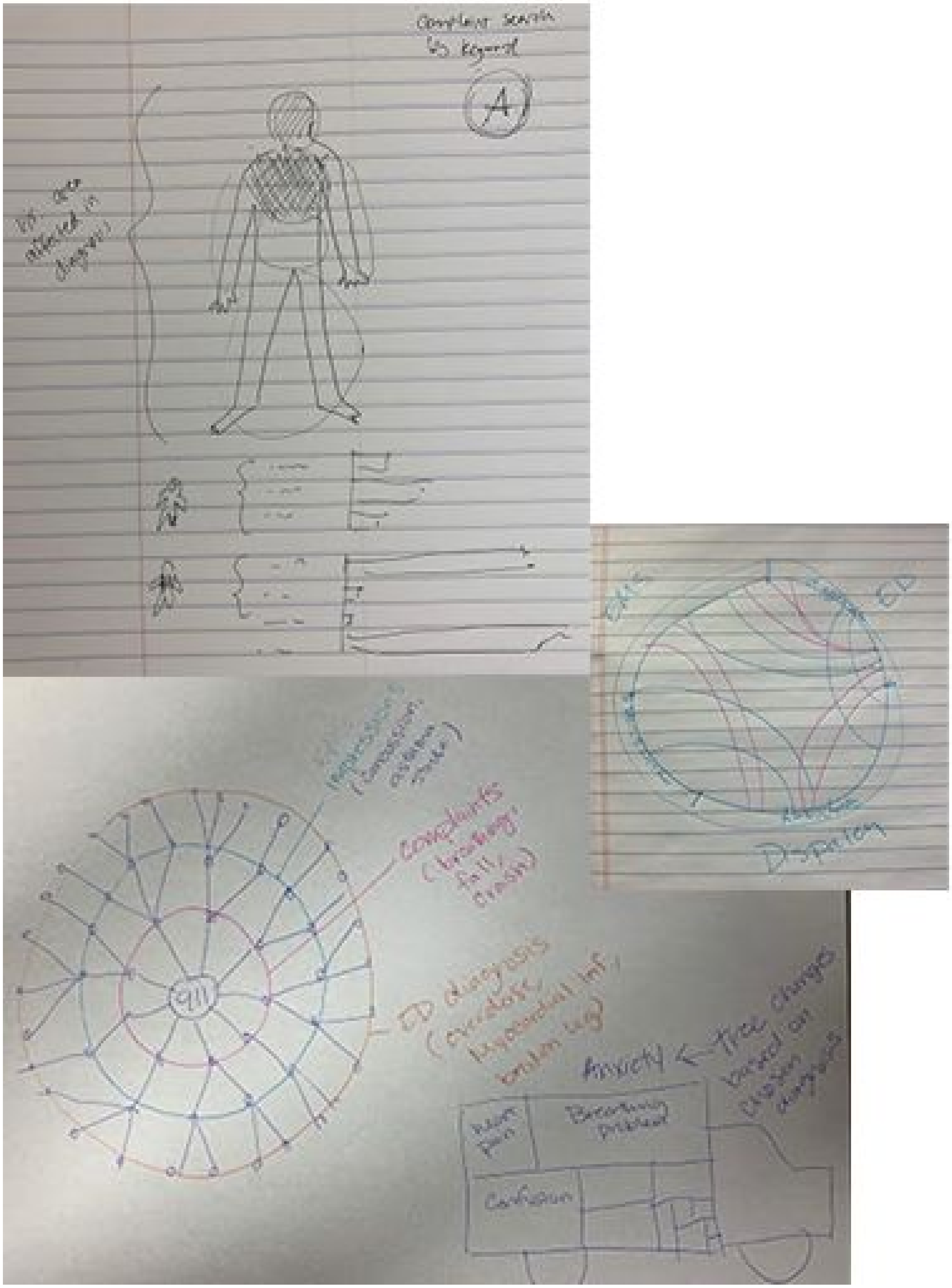
While the encoding matches the data, more coding could be introduced to highlight more features of the data. Specifically, semantic zooming and more use of color. There is currently no animation, but could be added through zoom and/or collapsing branches. The multiple views are coordinated and describe the top-down, and then bottom-up, organization of the data. Because there is only one year of data available, adding an animated time dimension is not possible. We should plan for animation as an optional feature.

The feedback received from Michael and Fangfei was both fair and helpful. They focused on what was going right and possible improvements. They had some fresh ideas that will improve the visual presentation of the data.

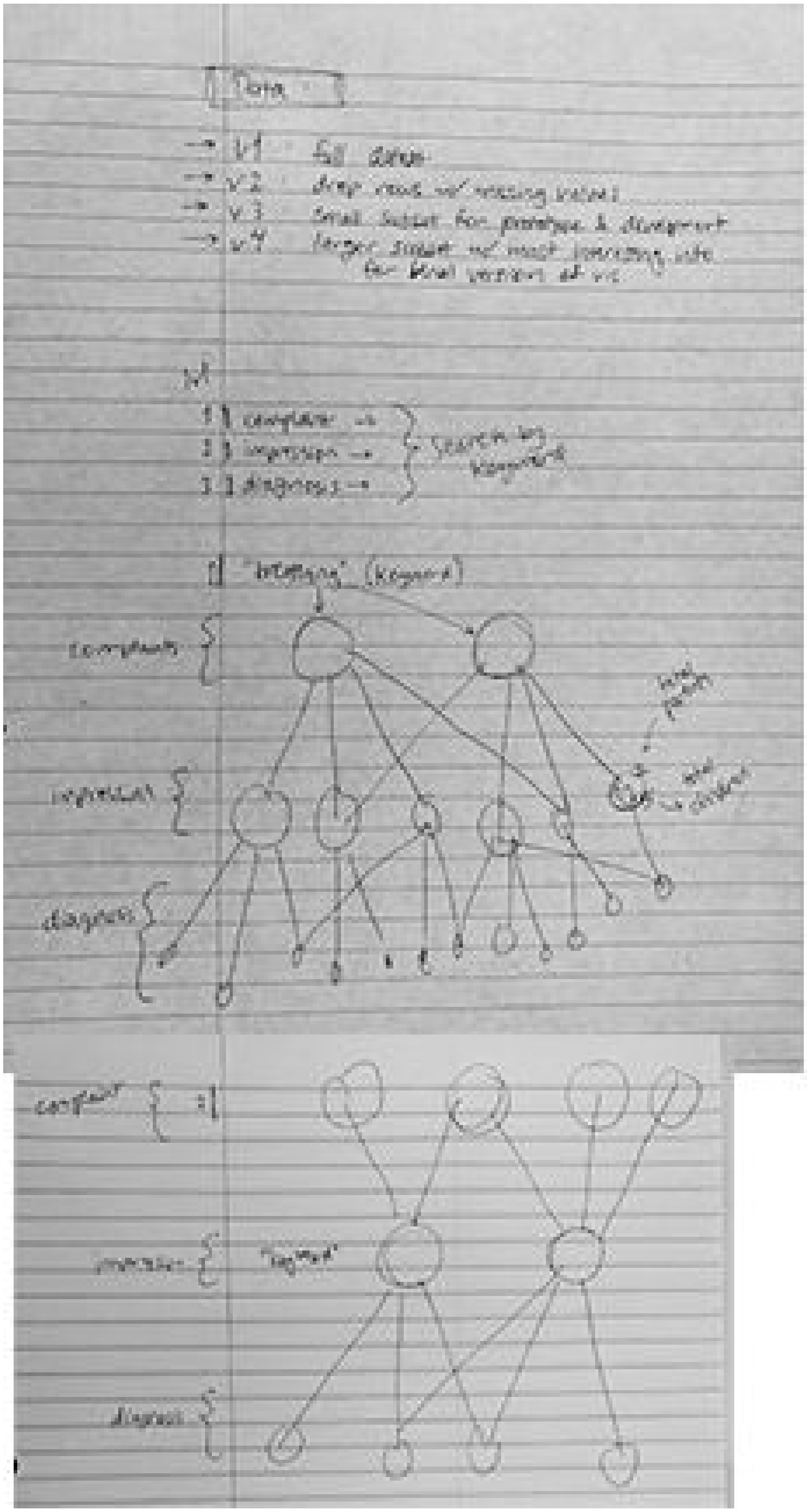
Video

https://drive.google.com/file/d/1gpg-a8g7EQu9rblPkPkbk8AeZsyWvMnI/view

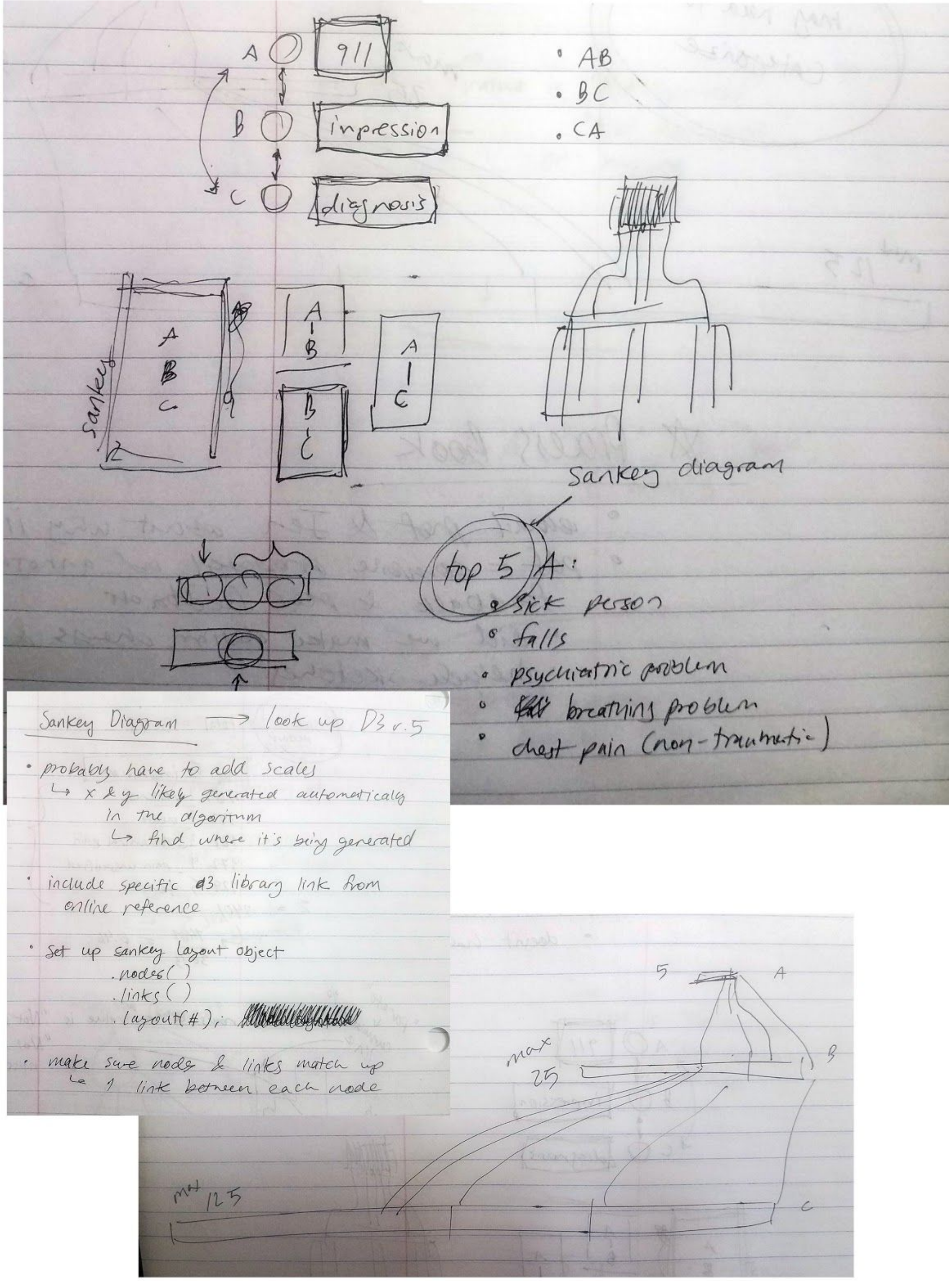
Brainstorming: We didn't end up using these designs, but they helped us think about the data.


Complaint search by keyword
A
ED
Dispatch
Complaints (breathing, fall, crash)
ED diagnosis (overdose, myocardial inf, broken leg)
911
Anxiety ← tree changes based on chosen diagnosis
heart pain
Breathing Problem
Confusion

Trees: An iteration on the tree from before, but this time we emphasize the flow structure.



Sankey: A suggestion from Jen to use the Sankey format to show our tree structure got us thinking about featuring the frequencies of events in our visualization.



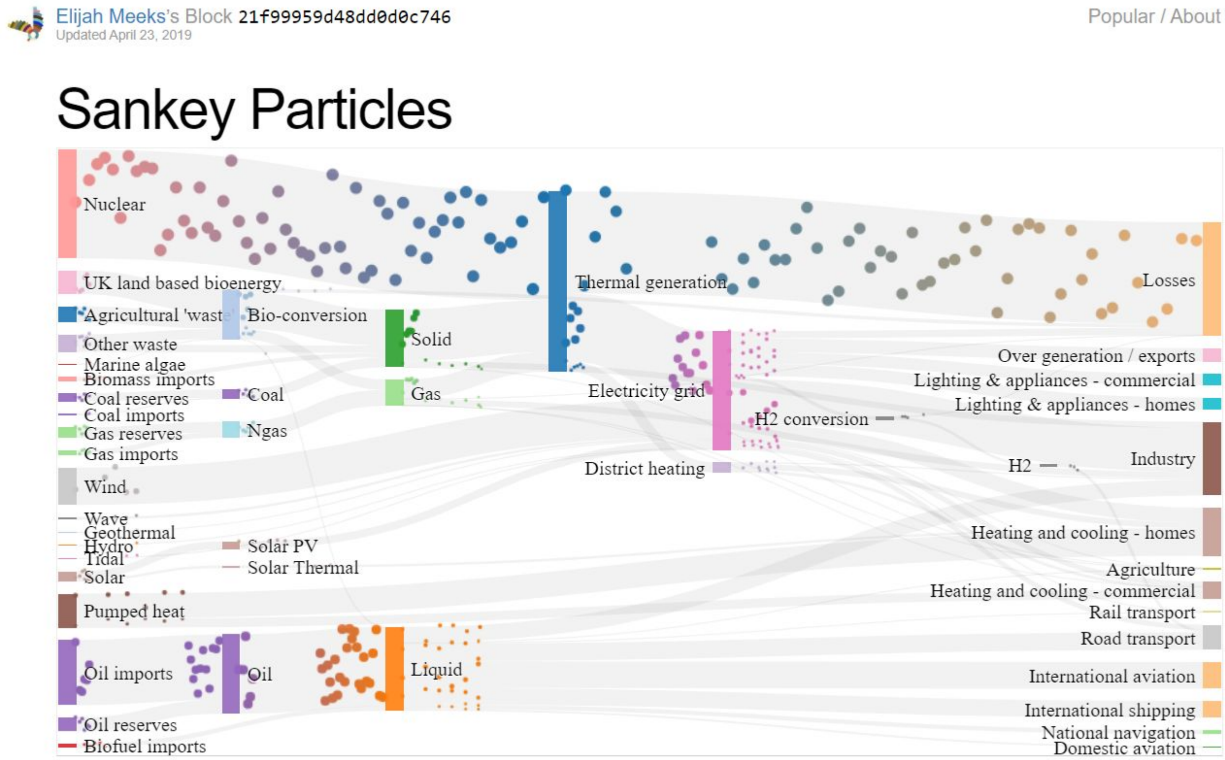
top 5 A:

- sick person
- falls
- psychiatric problem
- ~~fall~~ breathing problem
- chest pain (non-traumatic)

Sankey Diagram → look up D3 v.5

- probably have to add scales
  ↳ x & y likely generated automatically in the algorithm
    ↳ find where it's being generated
- include specific d3 library link from online reference
- set up sankey layout object
  .nodes( )
  .links( )
  .layout(#); [redacted]
- make sure nodes & links match up
  ↳ 1 link between each node

Sankey Reference: We looked at the reference Jen sent us and then did some more research on our own. This is when we really settled on Sankey as the primary view for showing the flow structure of our data.

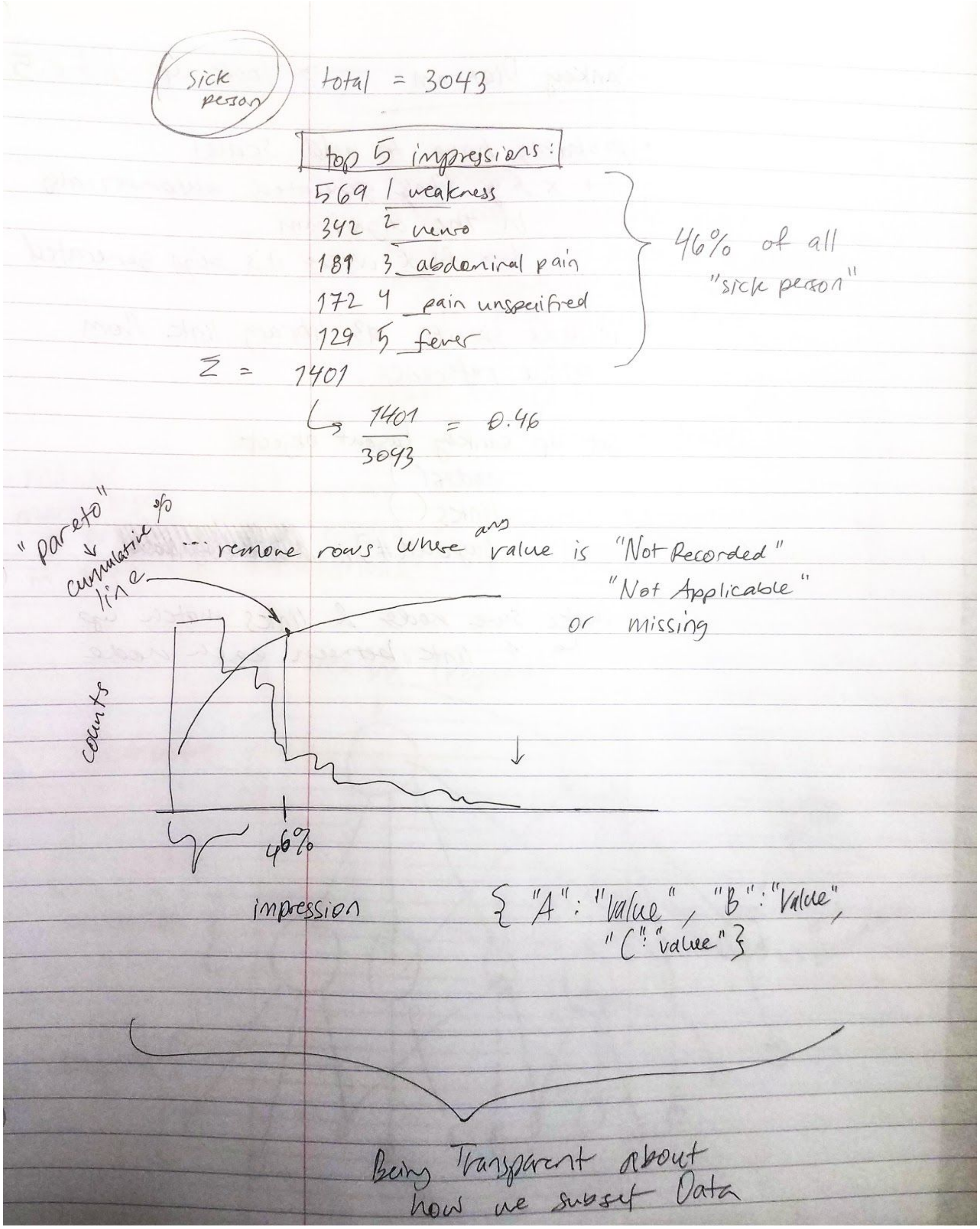


Elijah Meeks's Block `21f99959d48dd0d0c746`
Updated April 23, 2019

Popular / About

Sankey Particles

Bar Chart with Pareto Line: After deciding on the Sankey, we worked on getting a representative slice of the data that was small enough to show in that format. Leaving out all the extra information seemed like a waste though, and we decided to use a bar chart with pareto line to show the drop-off rates in frequency for different aspects of the data. Our initial calculations showed that we could capture nearly 50% of the data in the top 5 complaints, and we decided to go from there.



Sick peson

total = 3043

top 5 impressions:

569	1	weakness
342	2	neuro
189	3	abdominal pain
172	4	pain unspeified
129	5	fever

46% of all "sick peson"

$\Sigma = 1401$

$\hookrightarrow \frac{1401}{3043} = 0.46$

"pareto"
↓
cumulative % line

... remove rows where any value is "Not Recorded"
"Not Applicable"
or missing

counts
46%
impression

{ "A": "value", "B": "value",
"C": "value" }

Being Transparent about how we subset Data