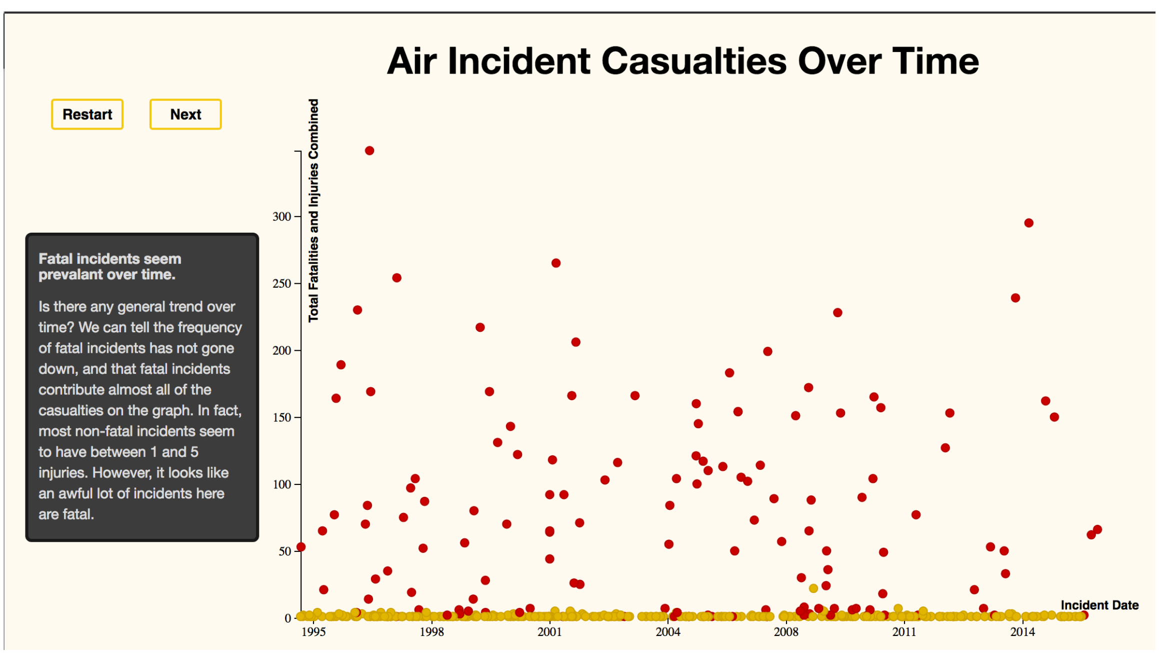
Team member names: Alex Herbig (solo project)
Dataset chosen: aircraft_incidents.csv
Analytical Tasks:

- Overview
- Details on demand (hovering)
- Filtering (done by storytelling style, not user)
- Sorting (done by storytelling style, not user)

Analytic Questions:

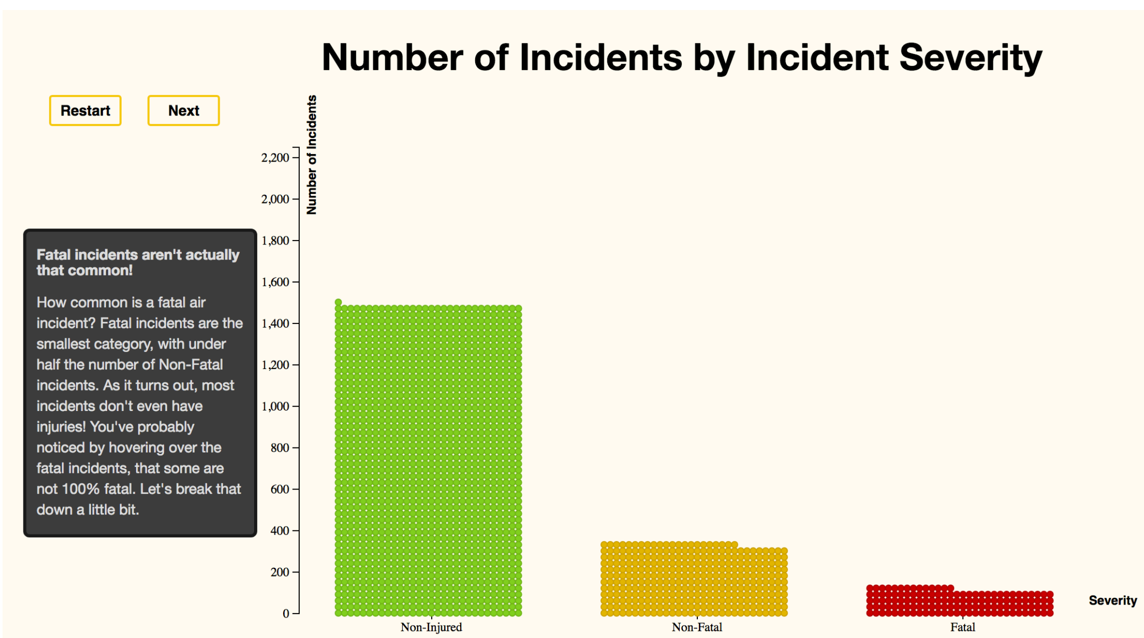
- Compare total casualties from different incidents with respect to time (vis 1)
- Compare total casualties in fatal vs. non-fatal incidents (vis 1)
- Compare number of incident types in the dataset (vis 2)
- Explore breakdown of fatal incidents with respect to time (vis 3)
- Which fatal incidents have higher non-injured rates, injury rates, and fatality rates (vis 3)
- Which fatal incidents involved lots of people? Which involved very few people? (vis 3)
- What is the distribution of percentage of injured vs. non-injured people in non-fatal incidents? (vis 4)
- How many people are involved in a typical high injury percentage flight? (vis 4)
- If there are a large number of people injured, is that likely to be a high or low percentage of the people on the flight (assuming non-fatal). (vis 4)

Quick note: the next button is used to transition from vis to vis. The restart button restarts the vis sequence at visualization 1. The page is meant to be progressed through one vis at a time.

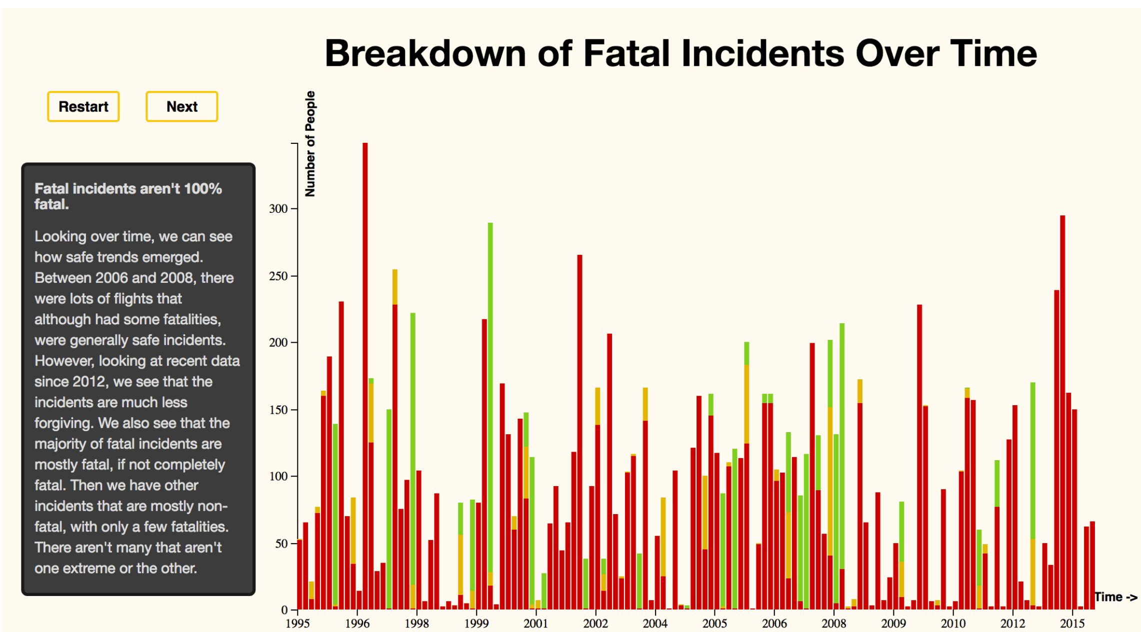


The first visualization of the project (seen above) shows a pretty overviewed view of the data. The data is presented with date on the x axis and number of casualties on the y axis. Number of casualties is calculated by adding up the number of injured people and number of killed people on each flight. The data points themselves are color coded to show the difference between fatal incidents (red), and non-fatal incidents (yellow). In my visualizations, I considered an incident as non-fatal if there were no deaths and at least one injury. Non-injured incidents are incidents which had 0 fatalities and 0 injuries. This is slightly different from how the original dataset categorized the incident severity. The first visualization can answer questions like, "Are there any general trends in the number of casualties over time?". "Which type of incident contributes the most to the total casualty count?". Users can hover over any of the data points to get more details about any of the incidents.

The second visualization shows an even more zoomed out view with more incidents. This is the visualization that brings in non-injured incidents (green). I decided to leave them off of the previous graph because non-injured incidents are not going to contribute anything except clutter to a graph of casualties, since they have no casualties. This visualizations answers questions like "What is the most common type of incident?", "How much more common is a non-fatal incident than a fatal incident?", and "How many non-injured incidents are there?".




The third visualization goes more into detail and takes a different style of graph, a stacked bar chart. I decided to go with a stacked bar chart because fatal incidents have 3 categories of people (fatalities, injuries, and safe). You can't represent 3 characteristics like that as one percentage, but I still wanted to break down each incident. With the number of incidents being somewhat small, I decided that a stacked bar chart was reasonable. I chose to go with time for the x-axis, although it's not an even axis, since each bar takes up a set width. Ticks are on the axis to provide baseline years for what year of data is around that tick. I chose time so that the graph could better be used to look for trends in the data, ex: if one area of the graph looked more green or yellow than another area. This was supposed to take advantage of subconscious image processing for finding trends. Other things I considered sorting by was total head count per incident, which would have made a prettier visualization, but overall would have provided less analytical power. One thing it could have shown better is whether the larger incidents tend to be safer or tend to be more fatal (or if there's no trend). I decided to go with date in the end though, since I felt that it was a more powerful visualization. This visualization answers questions like "What trends have there been in fatal incidents over time in regards to fatality rate and injury rate?", "are incidents that are fatal generally completely fatal, or mostly safe? Are there in-betweens?", and "Which incidents have a large amount of people involved? A small amount of people involved?".




The fourth visualization goes back to the scatterplot-type visualization. I figured that it was easy to encode injury rate as a percentage, since it's only injured and non-injured people for each incident. The x-axis in this graph is also date. One thing would have been lost though from the previous visualization, which is number of people per incident, so encoded in the size of the point is the number of injured people. This is slightly different than the number of total people in the flight, but I liked the idea of encoding number of injured people, because I found myself immediately asking myself "Well which one of these incidents has the most number of injuries?", when all the points were the same size. The radius encoding immediately helped reveal the answer. This visualization also answers questions like "Are there trends in the number of injured people over time?", and "How are the percentages of injuries distributed? Are there lots at the extremes, or lots in the middle?”. This visualization can be seen below on the next page.