

TaleSpin



2014

Visualizing airplane accidents over time- Hassan Jannah, Siddharth Agrawal, Sufia Siddiqui

TaleSpin is a data driven visualization of airplane accidents from 1919 to 2014 designed for anyone interested in the exploration of this topic. In this paper we review the design process, the tools used to gather and visualize the data and draw upon insights from usability testing.

Project Goals

Currently, a variety of online resources exist that have vast amount of data about airline accidents. However, most of the data is in tabular form and there isn't an easy way to visualize airplane accidents and different factors contributing to them in entirety. Our visualization tool aimed to:

1. Uncover myths about airplane accidents
2. Uncover stories behind airplane crashes
3. Explore the relationship between different attributes of the data

Related Work

A variety of prior works influenced our visualization both in terms of design and how to communicate our story effectively. Below is a discussion of these works and how they influenced our design.

1. *US Gun Deaths* (<http://guns.periscopic.com/?year=2013>)

US Gun Deaths is a compelling visualization of all the gun deaths that occurred in the United States in 2011 and 2013. It showcases the gravity of issue by visualizing the number of deaths and how long the person would have been able to survive. Although our dataset is different, this visualization influenced our design in three ways: (1) it helped us to understand how to show a variety of events taking place, (2) effective ways in which to filter the visualization, and (3) the effectiveness of multiple facets within a visualization. This can be seen in our initial design, which had a single line for each plane that resulted in an accident and in our final design where we had multiple facets within our visualization such as a stacked bar that was cumulative count of survivors and fatalities in an airline crash and a line chart that was drawn year by year as the visualization progressed with time.

2. *Out of Sight, Out of Mind – Drones* (<http://drones.pitchinteractive.com/>)

The drones' visualization showcased the casualties of drone attacks in Pakistan from 2004 - 2013. This visualization showed us how effective an introduction could be to set the stage for the visualization and multiple aspects of the visualization were reflected in our own work. For instance, the dropping of the planes and gradual increase of the bar charts was an influence in our design as was the stacked bar which kept a cumulative count of the survivors and fatalities due to the crash. Our visualization also wanted to expose narratives of airplane crashes so people could observe stories under the accidents just like the news section of the drones' visualization.

3. *Flight Quest – GE Data Visualization*

[\(http://visualization.qeblogs.com/visualization/flight-quest/\)](http://visualization.qeblogs.com/visualization/flight-quest/)

The flight quest visualization is a visualization of the cost and timesaving's by minimizing two and a half minutes of flight time per medium sized plane every day. The minimalistic design of this site and the storyboard influenced our final storyboard and color scheme.

4. *Reduce your odds of dying in a plane crash - A pre-flight check*

[\(http://www.informationisbeautiful.net/visualizations/reduce-your-chances-of-dying-in-a-plane-crash/\)](http://www.informationisbeautiful.net/visualizations/reduce-your-chances-of-dying-in-a-plane-crash/)

This infographic has used several datasets available over the internet to display the airplane accidents based on country, aircraft type, month of flight, seating plan etc. It helped us to think beyond our underlying question of visualizing accidents in different phases. Based on that we decided to integrate 100 worst accidents and 100 unusual accidents in our visualization. While the infographic does a good job at visualizing the different aspects of airplane accidents, its static nature limits the user to explore the data. To overcome this limitation, we decided to make our visualization dynamic, and also added the Sankey diagram to let user explore the data in detail.

Visualization Description

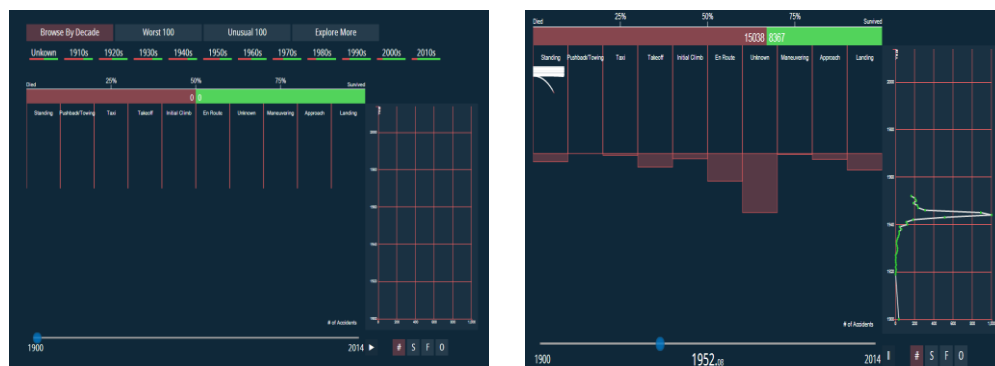
We designed the web visualization in way that support a certain flow. First we wanted to introduce the content of visualization, frame it, and then present some of the primary questions it answers. The main animation will allow the users to play the story that we want to tell and explore it visually. In supplementary visualizations, we present two lists of narratives related to airplane accidents that might be interesting to the public. Finally, we wanted to hand over the control to the user to explore the data set further. Using the sankey diagram, the plan was to allow the users to interactively choose different data dimensions (e.g. decade, airplane make, flight nature etc.) and see the relationship between those dimensions and the survival rate of passengers.

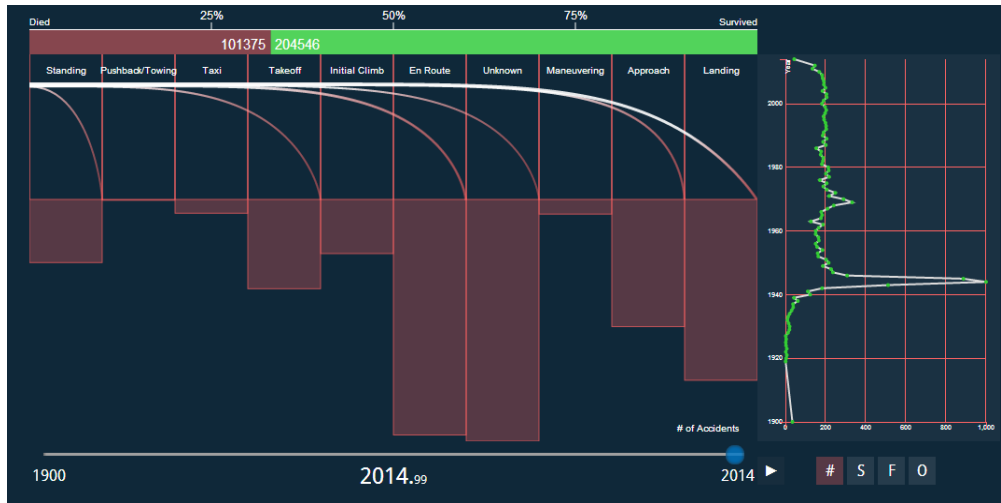
Part 1: Introductory Storyboard



Part 2: Main Visualization

From the start, we wanted to have all the visual elements visible on a single screen with not page navigation. The page would be like a movie player with controls at the top and the bottom and the story sandwiched in the middle.





Layout elements:

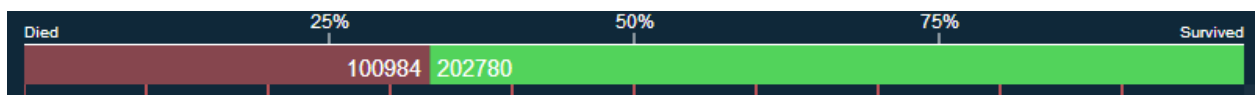
1. Tab options:

There are four different aspects to our visualization, which we have tried to explore and the user can click on the different tabs to explore the dataset.



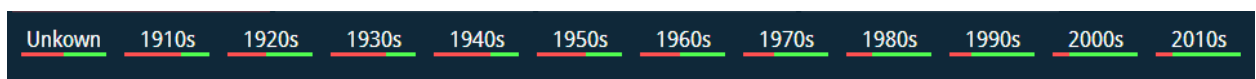
2. Survived/Died stacked bar chart

The stacked bar at the top of the animated chart provides a cumulative count of victims and survivors of plane accidents. It is a dashboard like visual that shows both the numbers and percentages of both groups and how those numbers change over time.



3. Decade filtering and information preview

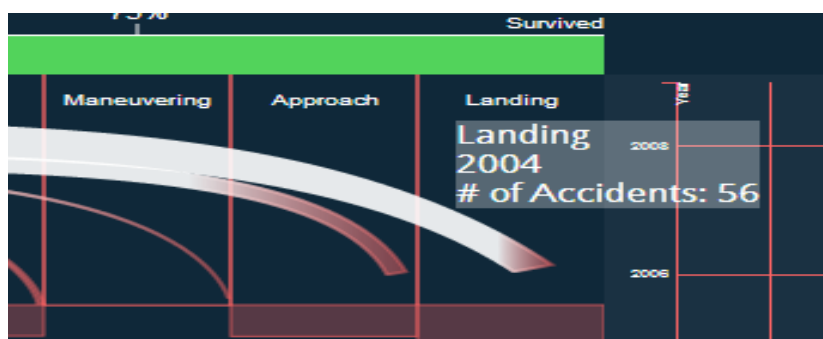
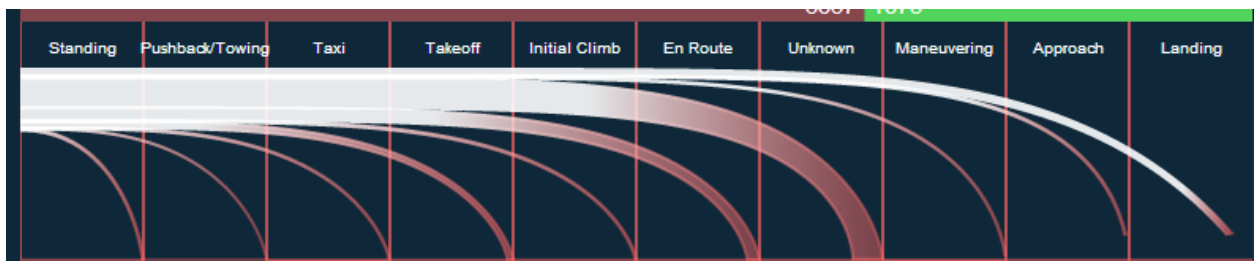
Due to the large dataset size, we wanted to provide a filtering mechanism for the users. However, we included a stacked bar chart at the bottom of each decade to give a preview of the data in that decade. It also shows clearly how the survival rate has increased significantly over time. We think that this visual is also intuitive because it follows the same structure and style of the main survival stacked bar. This element became functional, aesthetic, and informative at the same time.



4. Animated Falling phases lines

The falling lines section with different phases that resemble a 'faucet' feeding bottom bars

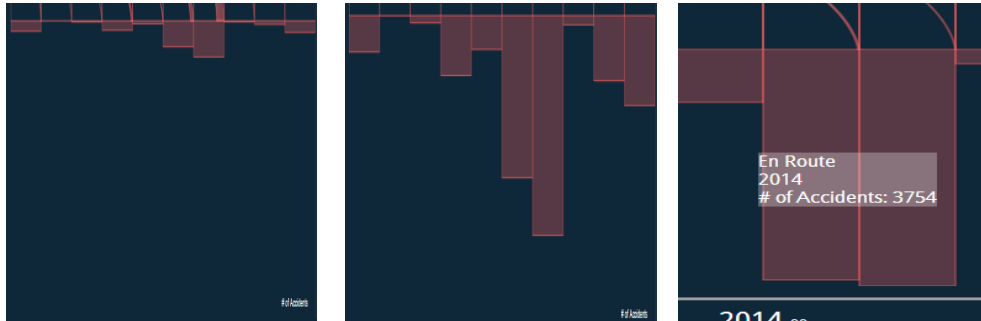
- The thickness of each line represents the number of flights at the year being animated. The line animates from left to right going through each phase of the flight. As the line passes through a phase, a portion of the line splits and starts falling to the bottom of the chart. The line also changes colors slightly to show highlight this split. The size of the portion represents the number of accidents in that give phase. The downward-curved shape of the line simulates falling. It also can be seen as a 'faucet' feeding the bottom bars, which grows overtime. The size of the curve and the increase in bar size are not directly proportional.
- Hovering on any lines reveals the actual information it represents (year, phase, and number of accidents).
- To draw each curve, we created two almost identical Bezier curves, shifted one of them downward based on the thickness, and then joined the edges programmatically give the illusion of a full shape. We didn't use a single Bezier with varying thickness because lines cannot have gradient colors.
- This element was not always obvious to users and a small axis on the left could help make it more intuitive.



5. Bottom bars cumulative over time

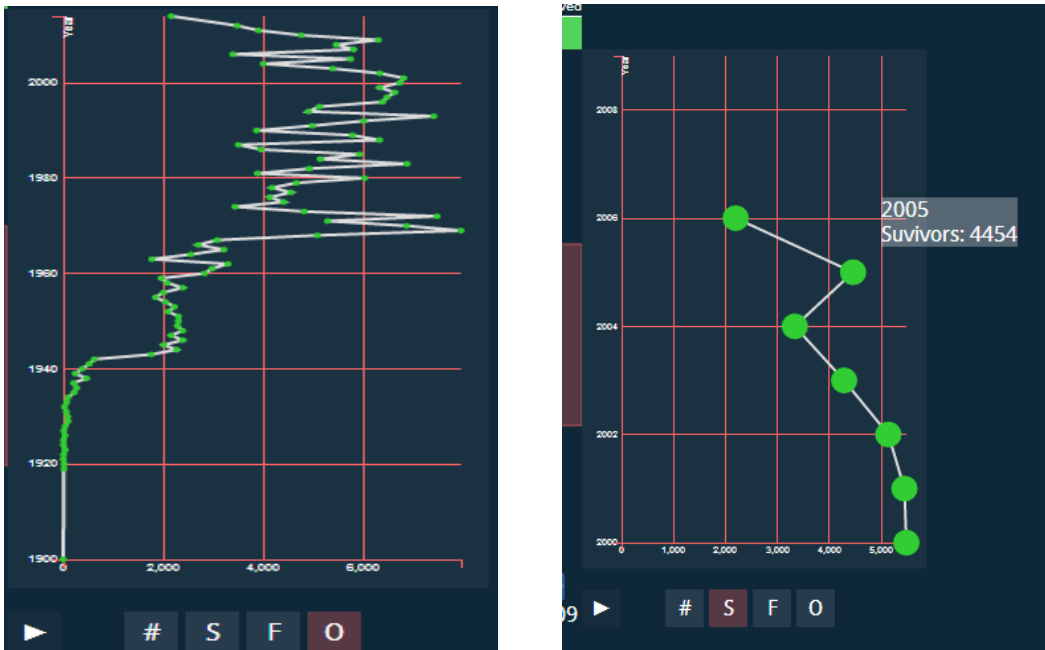
This visual helps show when accidents happen over the entire period of time. By the end of the animation, this chart helps show that most accidents happen in the first 3 minutes (phases 1-3) and the last 8 minutes (last 3 phases) combined. It also shows that there are a large number of accidents, which we don't know when they happen.

Hovering over the line reveals its underlying data (phase and selected measure: # of accidents, survivors, fatalities, or occupants).



6. Side Time-series Chart

The side time series line chart represents a breakdown of the bar chart by time instead of by phase. The choice to put it on the side was for layout purposes. The reason it grows vertically from the bottom instead of horizontally on the left is because the portrait-like dimensions of the chart allows more vertical space for the large dataset. A horizontal time series chart would be too squeezed. Some users suggested changing the layout of the element to be horizontal below the bar charts, which is a great suggestion for the future.

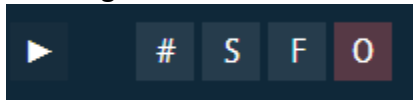


7. Play-button and slider to allow users to control browsing

This allows the user to sit and watch the animation or jump to any specific point in time. The animation accelerates gradually overtime to avoid boring the users with repetitive actions. It also gives a dramatic effect to the animation and allows the users to jump faster to final phase to see the whole picture.

8. Data change buttons

The data represented in the bottom bar chart and the time series can be changed dynamically at anytime to show the number of accidents (default), survivors, fatalities, and occupants. Unintentionally, The button set reads '#SFO' which some users found amusing.



Part 3. Narratives (Worst Airline Accidents and Unusual Stories)

The third part of the project involved showing two lists of individual special flight records:

1. Worst flight accidents based on the number of total fatalities order by severity.
2. A list unusual accidents obtained from a website ordered by time.

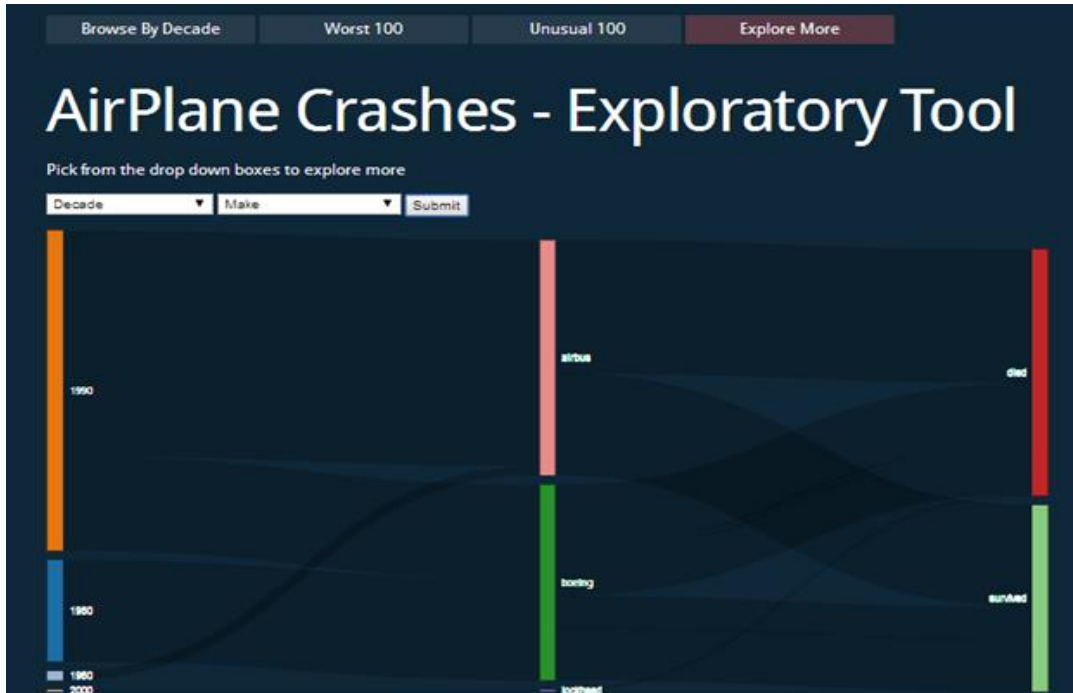
Each item on the list shows the date, the country location of the accidents and the number of occupants and fatalities. A user can click on any time to view full details of the accident including a narrative description of what transpired in it.

We didn't have enough time to represent this list visually. One idea was to represent the lists on XY scatters or a time series charts. If given time, we would like to include some pictures of these accidents if possible.

Browse By Decade	Worst 100	Unusual 100	Explore More
Tuesday 11 September 2001, United States of America O:92 (81 P, 11 C) F:92 (81 P, 11 C)	Tuesday 11 September 2001, United States of America Date: Tuesday 11 September 2001 Operator: American Airlines Type: Boeing 767-223ER Nature: Domestic Scheduled Passenger Departure: Boston-Logan International Airport, MA (BOS/KBOS), United States of America Destination: Los Angeles International Airport, CA (LAX/KLAX), United States of America Location: New York, NY (United States of America) Phase: En route Occupants: 92 (81 Passngers, 11 Crew) Fatalities: 92 (81 Passngers, 11 Crew) Survivors: 0 Ground Casualties: 1600 Airplane Damage: Destroyed Narrative: In a massive coordinated terrorist attack 19 men hijacked 4 jetliners on the		
Monday 12 August 1985, Japan O:524 (509 P, 15 C) F:520 (505 P, 15 C)			
Tuesday 12 November 1996, India O:312 (289 P, 23 C) F:312 (289 P, 23 C)			
Sunday 3 July 1988, Indian Ocean O:290 (274 P, 16 C) F:290 (274 P, 16 C)			
Wednesday 21 December 1988, United Kingdom O:259 (243 P, 16 C) F:259 (243 P, 16 C)			
Monday 12 November 2001, United States of America O:260 (251 P, 9 C) F:260 (251 P, 9 C)			

Part 4. Sankey Diagram

Sankey diagram serves as the exploratory tool for the visualization. The screenshot below shows a possible scenario where Decade is on the left, and Survival status of the passengers is on the right. Intermediate nodes group the Make of the airplane, and show the number of fatalities in a particular decade based on the make of the airplane. The thickness of the each link encodes the magnitude of information from source to target.



Initial Design

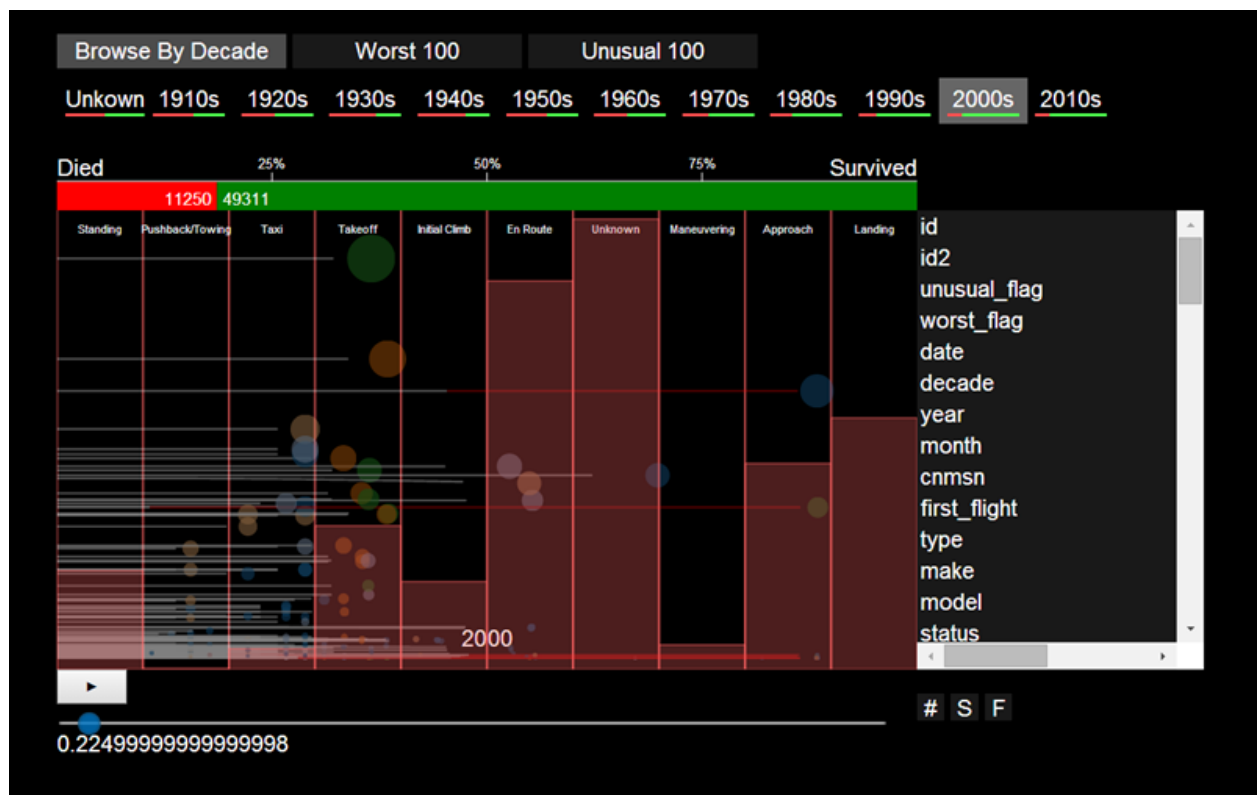
Color choices to give a radar-ish effect

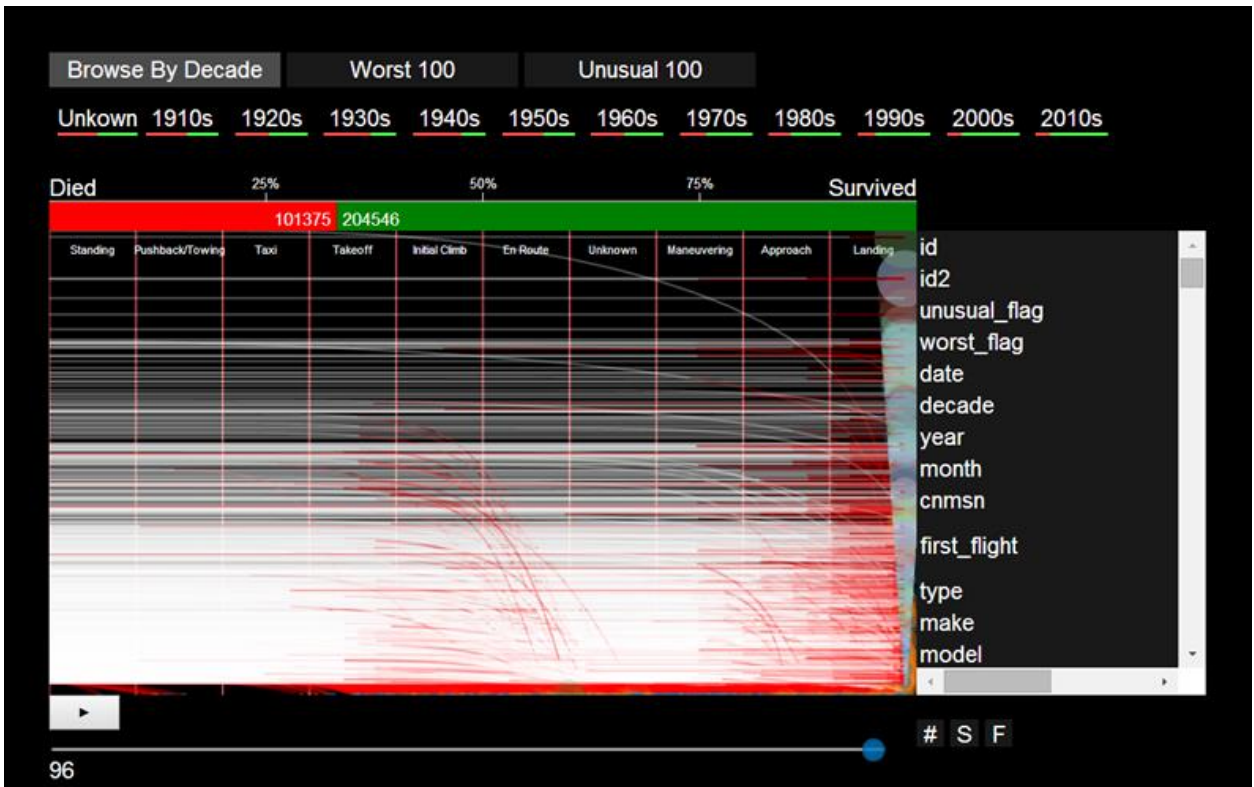
The color choices of green and red resemble a radar/sonar monitor often seen on movies. The colors also convey some meaning as they are used to trigger intuitive correlations of safety and danger. The choice of a dark background would make it easier to focus on the animated part and make them stand out more than on a white background. Based on the feedback from the users, we decided to change our design to the current one.

Old Design

Initially each individual flight was animated by a single line and animate in the following way:

- The flight line starts on a vertical axis representing the number of people on board
- The line changes color when the plane enters the phase in which it had an accident
- If no one survived, the plane will fall
- If some people survived, the line will dip to hit the opposite y-axis at the point that represents the number of survivors.
- The size of the circle of the flight also shows the number of occupants.
- The color of the circle represents the nature of the flight.





Data Source

The original NTSB data dump was full of missing values and inconsistencies. [The Aviation Safety Network](#) database provided a much cleaner and data set, which contains all the values, required for our targeted visualization elements. The only drawback was that the database was actually a set of tabular web pages. To prepare the data for the project, we used a combination of tools and techniques:

1. Using Beautiful Soup, we generated a list of the entire page URLs that needed to be scraped.
2. We used Kimono Labs web scraping tool in order to scrape the entire list of URLs
3. Used python to download the data from Kimono Labs because the API only retrieves 2500 records at a time. The script also converted the output from JSON to CSV
4. Used excel extensively to extract quantitative data values from the data (e.g. occupants and fatalities) and to clean up some inconsistencies in the numbers. We also used excel to extract additional textual information such as departure, arrival and accident location countries, and airplane manufacturers.

Midway through the project, we discovered that the scraped dataset was not complete and we had to repeat the above steps to get a complete dataset (16,284).

For the 100 worst accidents, we ranked the accidents by the number of air and ground fatalities. For the unusual lists, cross-matched our list with a list on [PlaneCrashesinfo.com](#).

Tools

1. Beautiful Soup - We used BS to generate a list of the entire page URLs that needed to be scraped from Aviation Safety Network database.
2. Kimono Labs API - We used this API to automate the scraping of entire airplane accident data from the list of URLs.
3. CSS webkit animation - We used HTML, JavaScript and CSS webkit animation to create the introductory storyboards.
4. D3 - We used the JavaScript library D3 to graph and animate our data and JavaScript/HTML/CSS for the rest. We also used code snippets from other D3 designs to help build the curved lines, namely the [Bezier curves](#) by Jason Davies. Most visual components were designed manually.

Project Steps

1. Data Research and acquisition
2. Data Curation
3. EDA with Tableau and Excel
4. Brainstorming and Storyboards
5. Initial implementation - phase 1
We are able to draw animated visual curves and simulate individual planes falling. Unfortunately, as the number of data objects increased, the browser was not able to render the visual elements in time (~50,000 by the end of the animation).
6. Redesign based on feedback and technical limitations.

Results and Usability Feedback

- Users liked our storyboards and it helped them to understand the context of the visualization.
- They liked decade-by-decade facet of the visualization where small green bars below every decade depicted the trend of number of survivors.
- As described earlier, #SFO tags were not intuitive on the main visualization page.
- They did not understand the purpose of width of the arc clearly. But once explained, it made sense to them, and they were able to relate it with the number of accidents in different phases.
- They liked the growing bar charts depicting the total number of accidents/survivors/fatalities/occupants with time.
- While bar charts helped to visualize the cumulative numbers, line chart was helpful to visualize the trend over the years. For example, one of the users pointed out that in 1944 there was a huge spike in total number of accidents due to Second World War.
- They liked the concept of Sankey diagram to explore the relationships between different attributes of the dataset.
- Overall, the visualization had multiple things to explore but without a guided tour it was complicated for users to figure out everything, which we wanted to show. A tutorial with notes would have been helpful for them to understand it more clearly.

Links

Demo - <http://people.ischool.berkeley.edu/~jannah/Planes/intro.html>

Github repository - <https://github.com/jannah/Planes/tree/new-design>

Contributions

Task	Hassan	Siddharth	Sufia
Idea and Research	30%	30%	40%
EDA		50%	50%
Storyboards	33%	33%	33%
Data Research	40%	30%	30%
Data Scraping	60%	40%	
Data Curation	60%	20%	20%
Intro		50%	50%
Web Layout	60%	20%	20%
Primary Animation	100%		
Sankey		50%	50%
Styling	60%		40%
Documentation	20%	40%	40%
Presentations			100%