FAA Wildlife Damage to Aircraft Lab 1-A

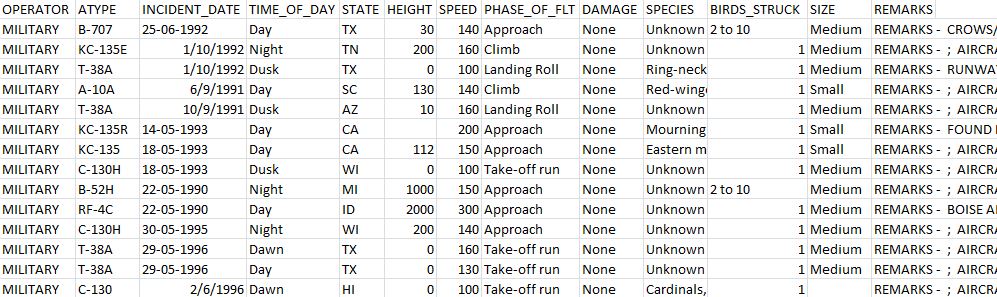
In 2010, as part of the VA research program, two students participated in VA internships at the Boeing Company. One intern worked with a Boeing Safety Engineer on the problem of bird strikes on commercial aircraft. All airframe manufacturers collaborate to study safety incidents to assure that air travel stays the safest mode of mass transit. In this way, manufacturers can change their designs and improve pilot training to minimize the impact of incidents like the Hudson River landing of US Airways Flight 1549 on January 15, 2009. Both engines shut down after hitting a flock of Canada Geese.

The visual analytics work done by the Boeing wildlife strike expert and the SFU VA intern\* led to four design change recommendations and a change to pilot training materials for all Boeing commercial aircraft.

\*That’s Andrew Wade!

Goal I: Let’s familiarize ourselves with the first step of visual analysis: ***understanding your data.***

**0. Before we start**

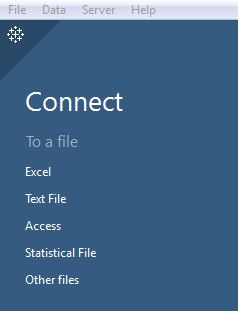
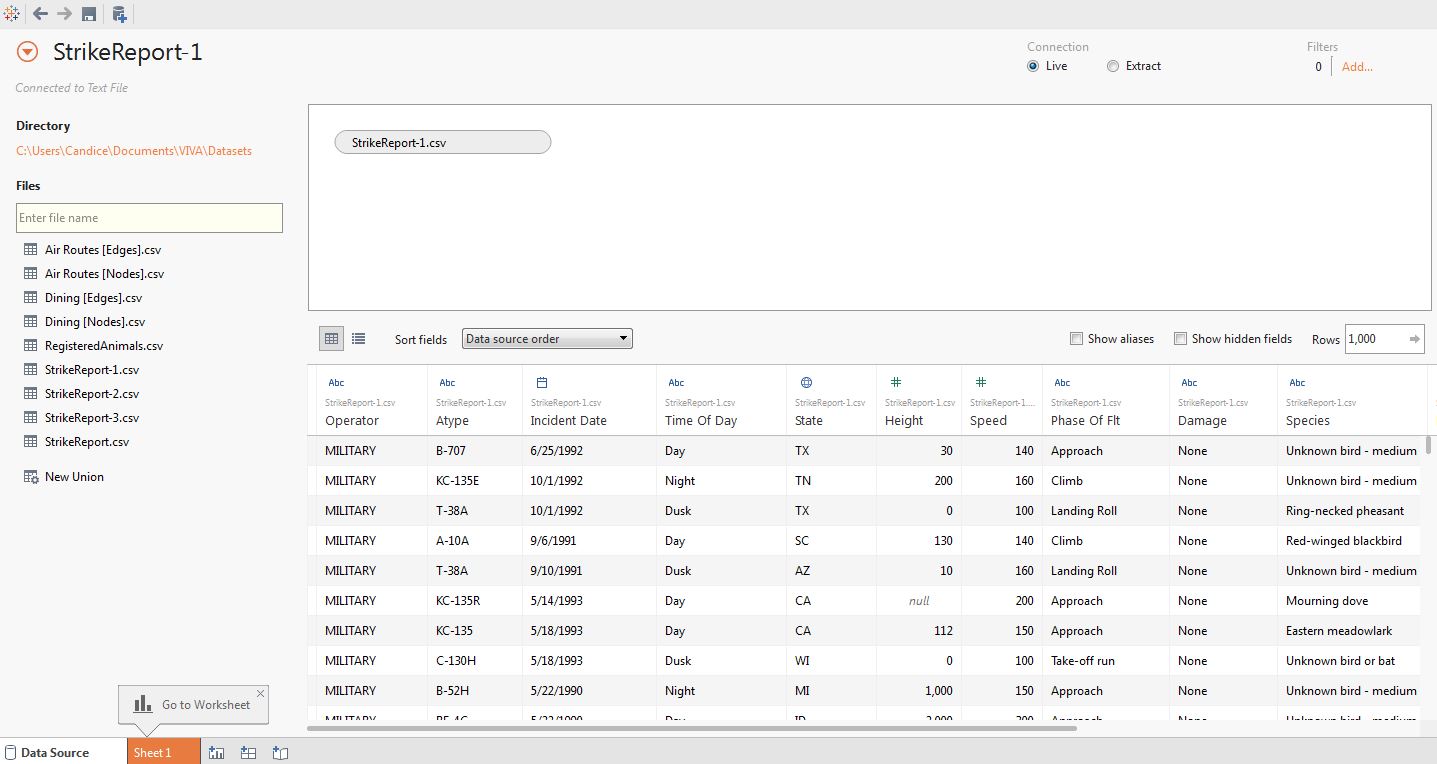
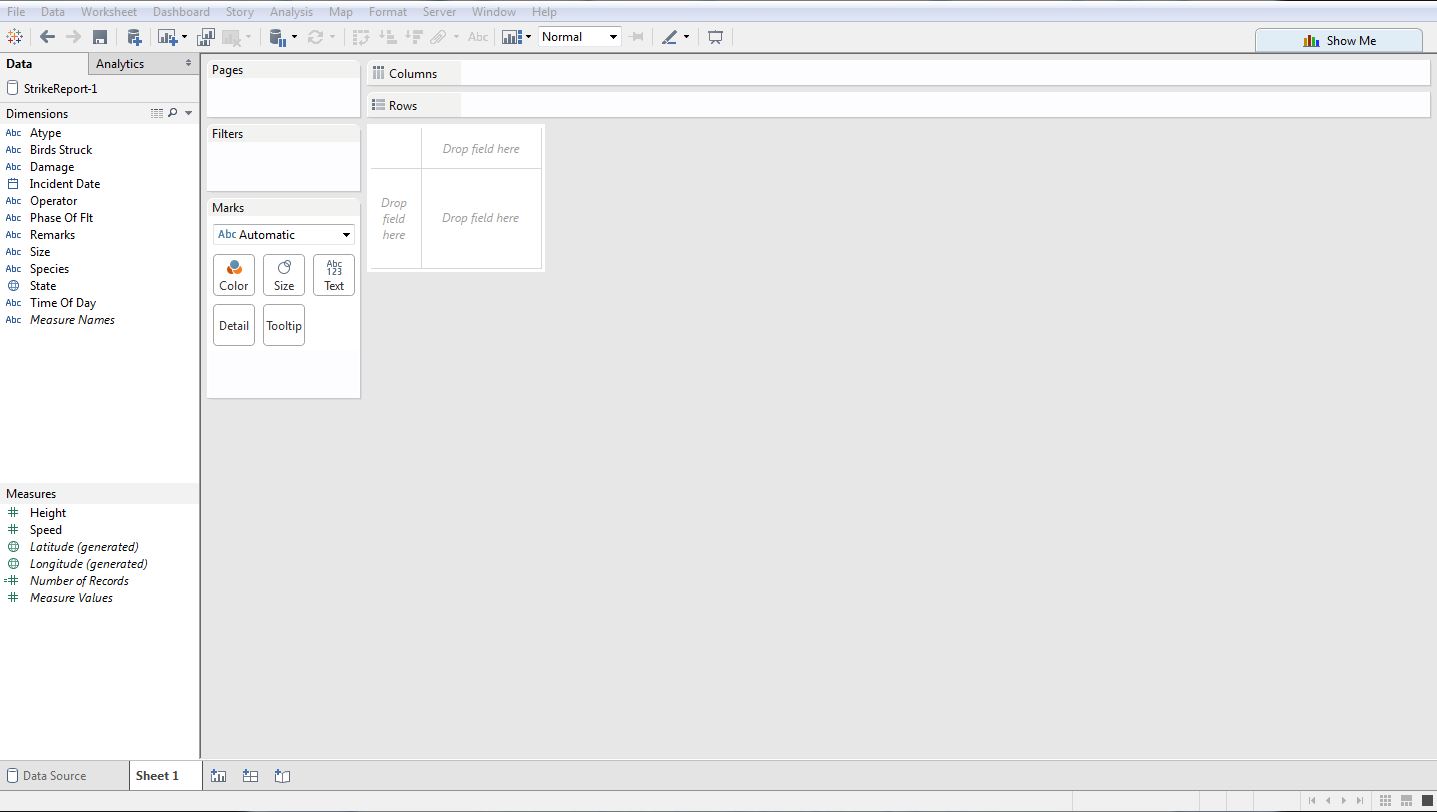
**StrikeReport-1.csv** is a simplified version of the full-fledged FAA dataset that Andrew used to analyze bird strikes at Boeing. If you open up StrikeReport-1.csv on Excel you will see something like this:  


You’ll see the columns:

* **OPERATOR** (airlines, military, private jets…)
* **ATYPE** – this is the aircraft type/model
* **INCIDENT\_DATE** – the date when something hit the aircraft
* **TIME\_OF\_DAY** (Day, night, dawn, dusk. Note: the length of dawn/dusk is much shorter than day/night)
* **STATE** (Note: this dataset has data points for places other than the US)
* **HEIGHT** (in feet) – how high the plane was
* **SPEED** (in mph) – how fast the plane was going
* **PHASE\_OF\_FLT** – “Phase of flight,” e.g. was it taking off? Was it landing?
* **DAMAGE** – how much damage there was on the plane
* **SPECIES** – what kind of animal hit the aircraft
* **BIRDS\_STRUCK** – the number of birds struck in one incident (you can run into just one bird or a huge flock of birds)
* **SIZE** – how big the animal is
* **REMARKS** – Notes taken by the pilot

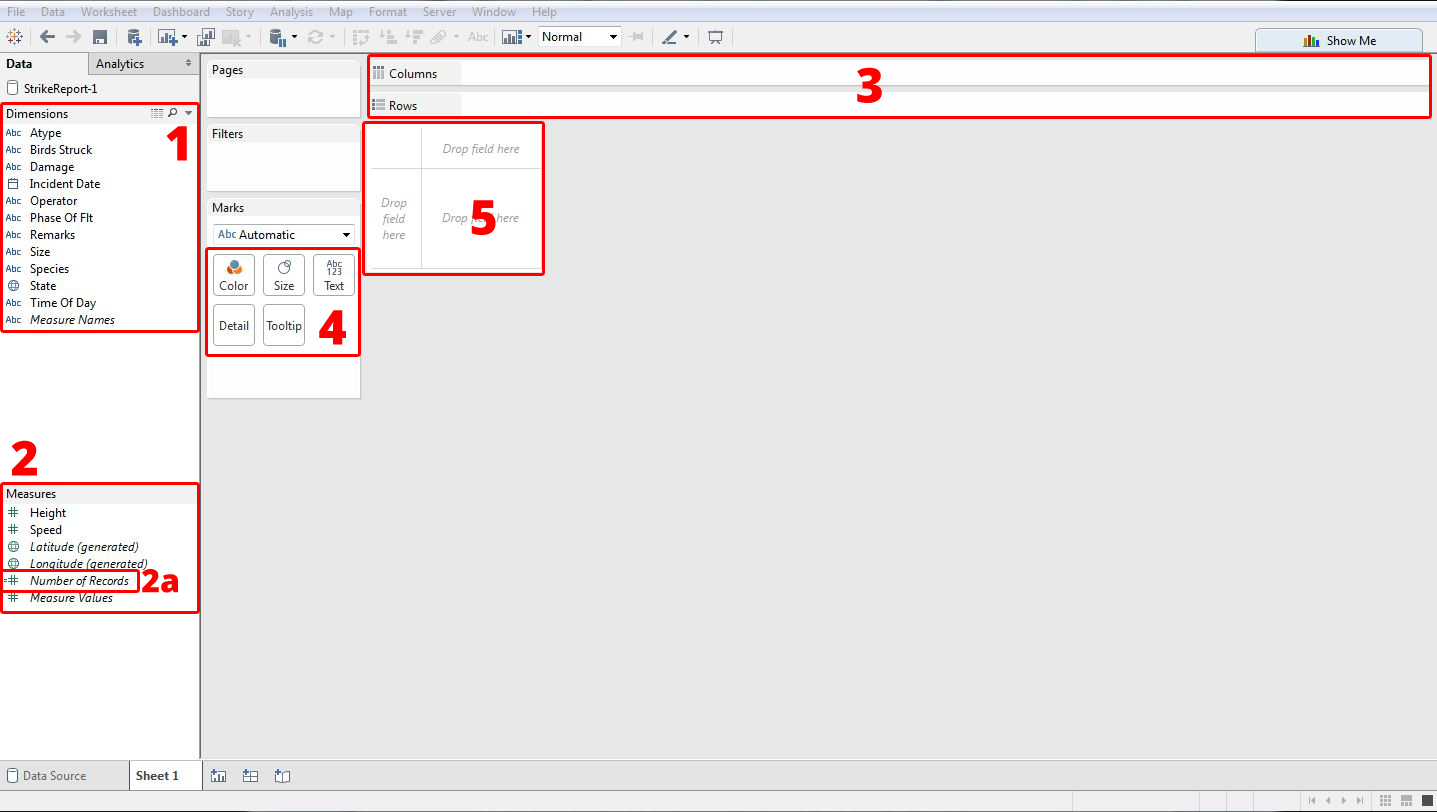
Now that we know what’s in the dataset, let’s import it into Tableau.

**1. Importing the file into Tableau**

1. Open the latest version of Tableau
2. When the main window opens up, you should see something like “Connect to Datasource”. In Tableau 9.3 it looks like this:  
     
   For our purposes - a **CSV file** is a **Text File** so click on Text File. Then find your file and open it up.
3. Now you should see something like this. The spreadsheet you see on the bottom right should look familiar (Operator, Atype, Incident Date, Time of Day…).  
   
4. When you look at the top right, it says Connection and the “Live” radio button is selected. This means Tableau is connecting to the data live, which means if you do data manipulations on Tableau it will affect your data (which could be risky!). Click on the Extract radio button.
5. Click on the orange button on the bottom left, called “Sheet 1”. If you make Tableau connect to an extract of your data, it will ask you to save your extract somewhere on your computer. Find a good place, and save it.
6. Now, that will bring you to the worksheet. Notice the column headers you saw in the table are now on the left side, under “Dimensions” and “Measures.” …Let’s go over some terms.  
   

**1.5. The Tableau lingo cheat sheet**

Here’s a quick guide for you so that you know what I’m talking about.



1. **Dimensions** are typically categorical data that can be used as independent variables. They are represented in **blue**.
2. **Measures** are typically numerical data that can be used as dependent variables. They are represented in **green**.
   1. In this walkthrough, **Number of Records** means the number of animals-hitting-airplanes incidents.

Caution: Tableau guesses what each column’s data type is (Is it a string? An integer? A float? A Boolean? A geolocation? A date?) and then guesses whether it’s a dimension / measure. It guesses, which means it can get it wrong. Always double check!

1. These things are **shelves** (because they sort of look like shelves) There is a column shelf and a row shelf.
2. These squares are called **marks cards** (because they sort of look like cards).
   1. But wait, what’s a **mark**? Think of it as a data point. The marks card controls how you see the data points.
3. The **canvas** is currently a blank table, but *when you drag and drop the dimensions/measures into the shelves and cards, visualizations take place here*.

When you drag and drop a dimension or measure into a shelf or a card, you will see it appear on the shelf or card as a “**pill**” (because they sort of look like pills).

Okay – so now we can talk about dragging and dropping *dimensions* and *measures* (e.g. Number of records) into *shelves* and *marks cards*, and we’ll see changes on the *canvas*. And if we want to tweak it, we can rearrange the *pills* on the shelves/marks cards.

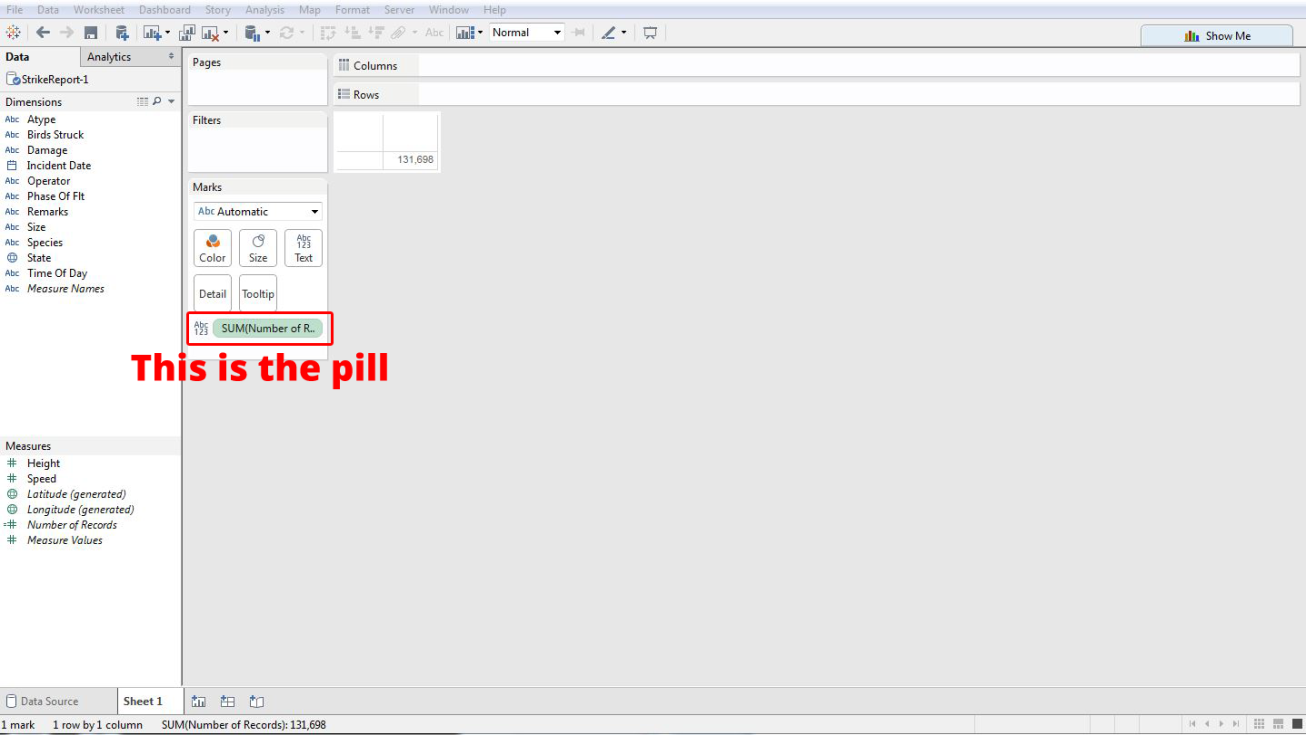
**2. Answering basic questions**

Now that we have our dataset in Tableau, we’re ready to do some more exploration about the dataset itself. Before we form any hypotheses, we need to know some general things like:

1. *How many incidents are reported in this dataset?*
2. *Over what years?*
3. *How many species are involved?*
4. *How is damage categorized?*
5. *How many operators are involved in strikes that resulted in aircraft getting destroyed or repair costs going over $2,000,000?*

We’ll answer these questions in this walkthrough.

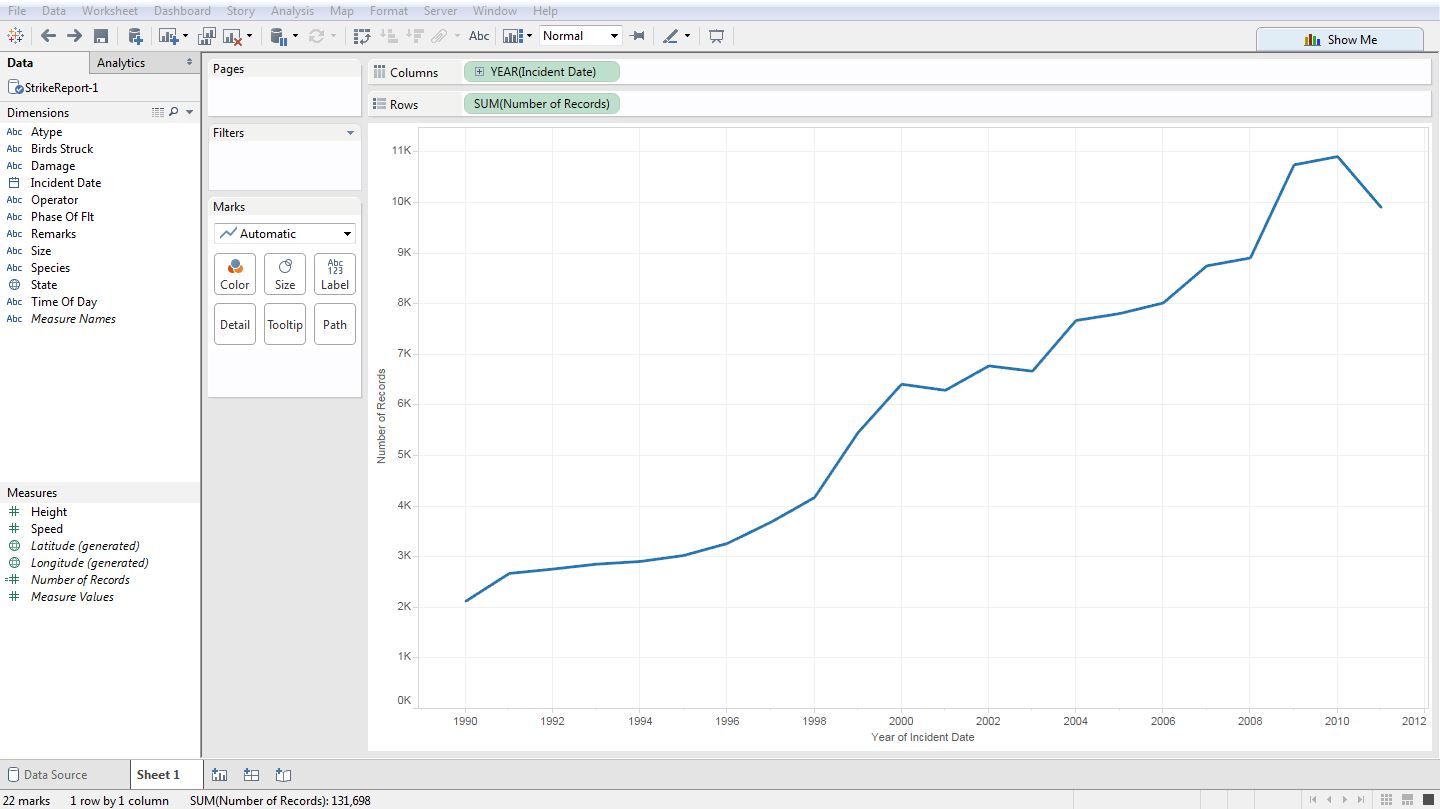
**i. How many incidents are reported in this dataset?**

1. In this lab’s context, the number of incidents is one of the measures: the **Number of Records**. Drag and drop that into the canvas (the big “Drop field here” square) and see what happens.
2. You probably got one number: **131,698**. That is the SUM of the number of records (wild life strike incidents), and you see it as a green pill under the marks cards.  
   

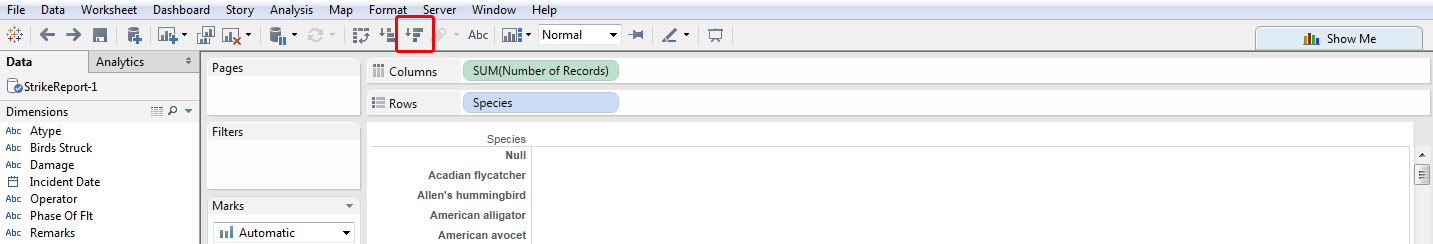
Right now you see a table with one number. It’s kind of boring, so let’s try something. On the top right corner, there is a button that says **Show Me** (with a colorful bar graph icon beside it). Click on that. When it expands, Tableau will tell you what kinds of visualizations it could make based on the stuff you put in the canvas/marks/shelves.  
Caution: Tableau’s Show Me feature is limited because Tableau doesn’t know your data very well (remember: it can only guess). So when you want to get creative, try experimenting with the “Marks” **drop-down menu** right above the Marks cards (by default it says “Automatic”).

Right now, Tableau doesn’t give you that many options. Also, since you only have one number (which is the sum of all incidents) it doesn’t really make sense to put it into a bar graph either. Let’s look at the number of incidents based on something. What about looking at the incidents over years?

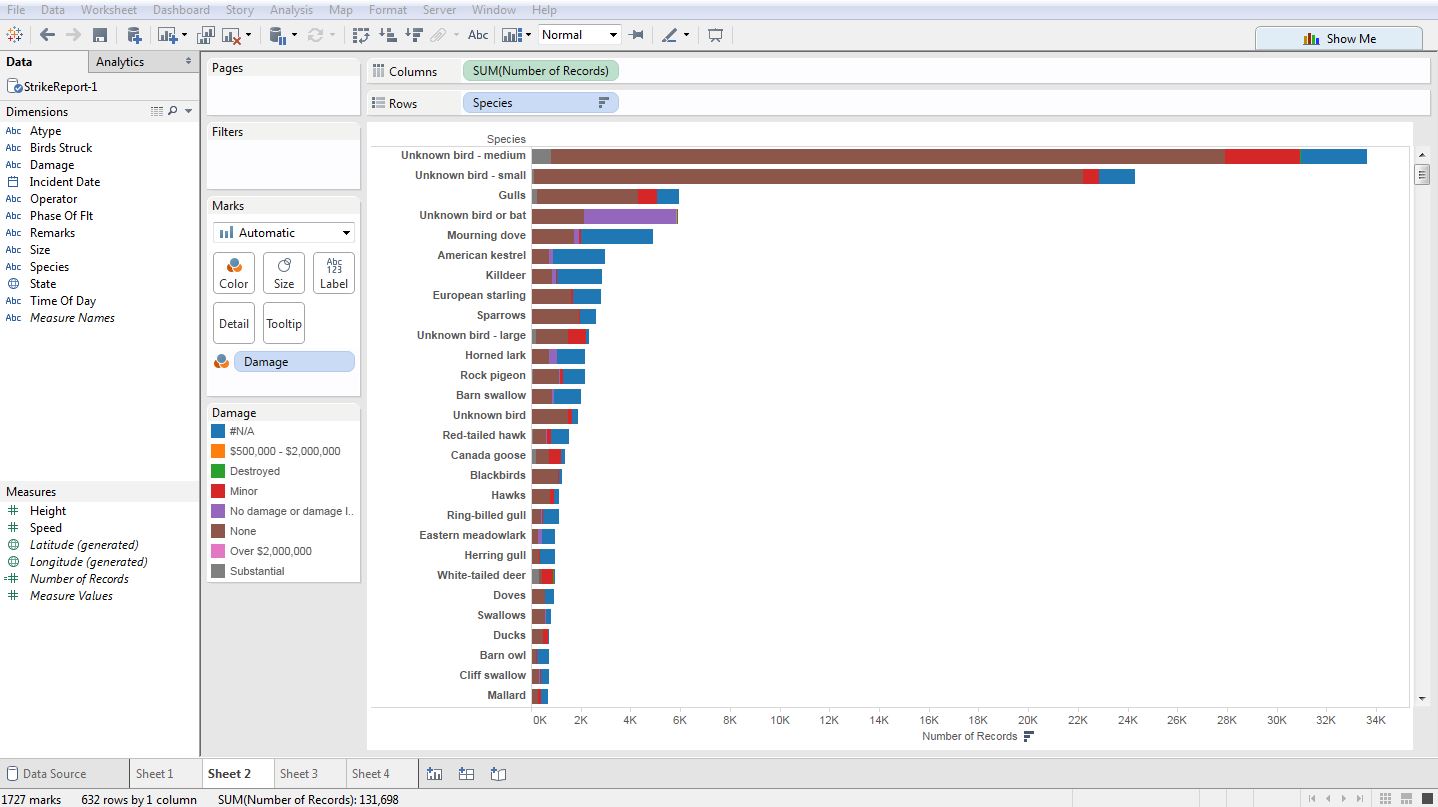
**ii. Over what years?**

1. To look at *how many* incidents occurred *when*, drag the dimension: **Incident Date** into either columns or rows. You should get a table that has the year and the number of records now.
2. Now open Show Me. We have lots of options! Feel free to experiment by clicking on each visualization icon that’s not half-transparent. Since we’re looking at change over time, and since we have data for every DAY, a visualization that would make sense is **lines (continuous)**. You should see something like this:   
   
3. Notice that the YEAR pill in the Columns shelf has a plus sign on it. If you click on that you can adjust the granularity of the data—from years 🡪 quarter 🡪 month 🡪 week 🡪 day. This act of zooming into the details is called **drilling down**.  
   Tip: Don’t like what you just did? Tableau has an un-do function (Ctrl or Command + Z)
4. Two questions for you:
   1. You’ll see that the records span from 1990 to 2011, and the number of records are increasing. Why do you think the number of records increased over the years?
   2. When you drill down into month – or even just the quarter, you’ll see something you didn’t see when you were just looking at years. What is it, and why do you think that’s the case?

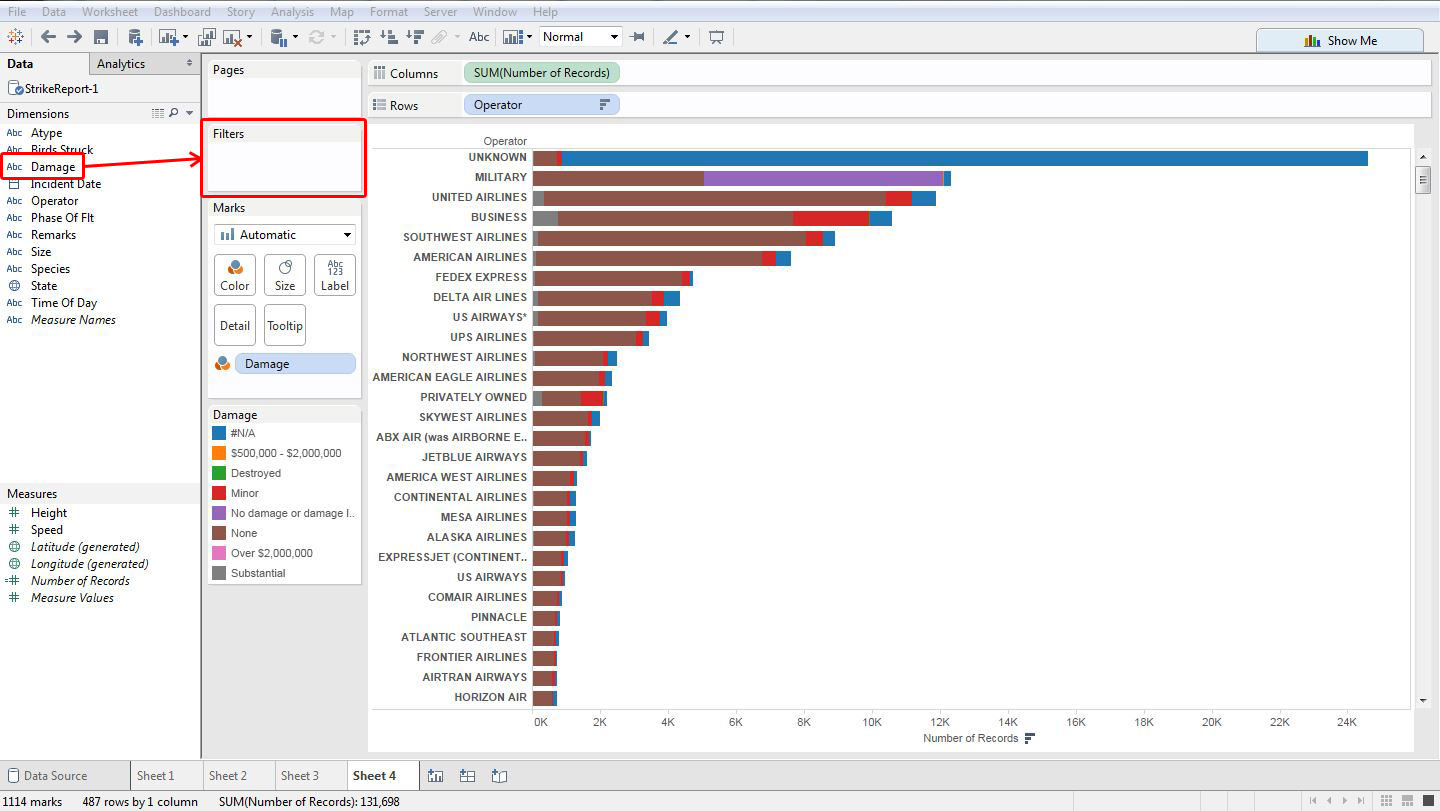
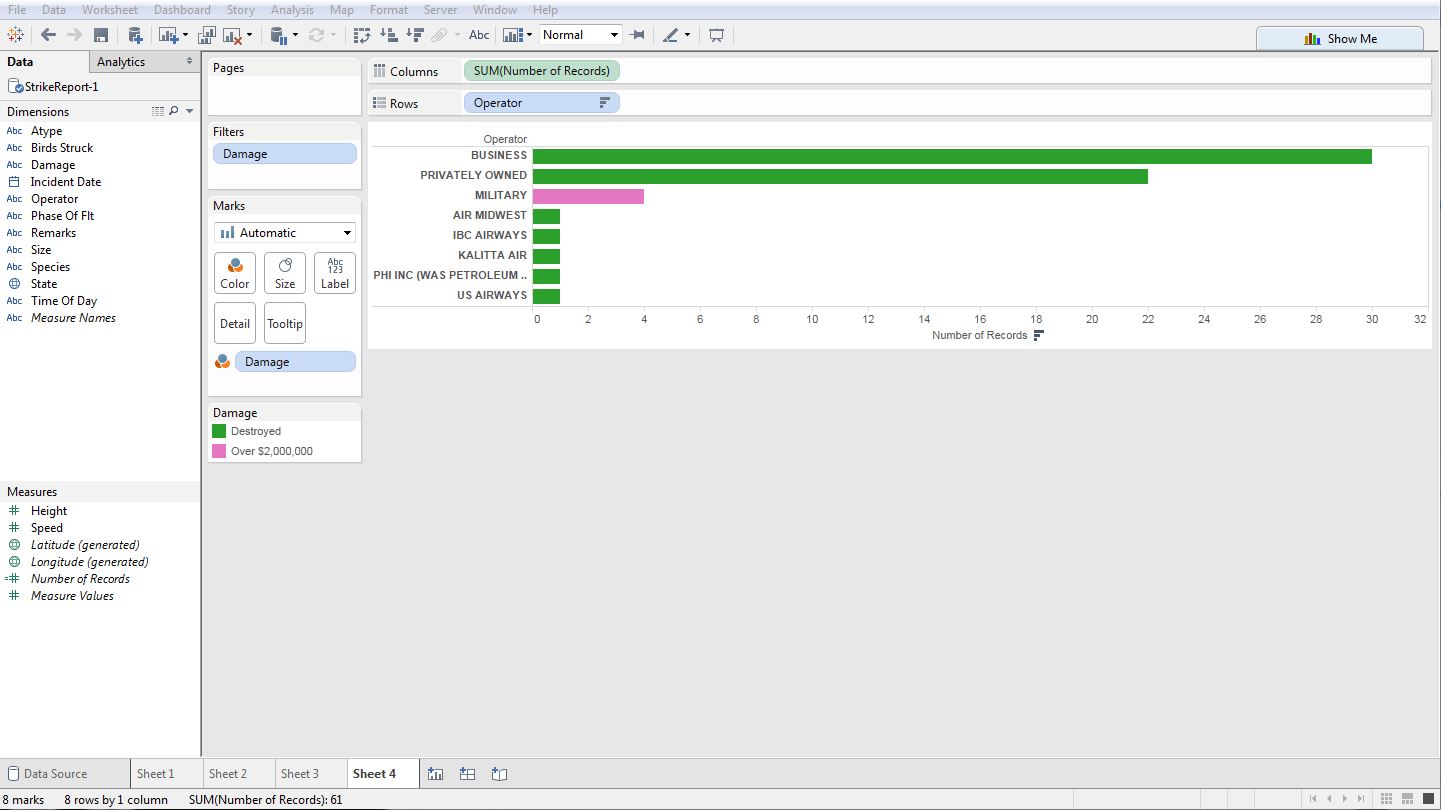
**iii. How many species are involved?**

1. Let’s create a new worksheet. We don’t want to just throw away the one we were just using. To create a new worksheet, go to the bottom left of the screen and click on the icon next to your Sheet 1 (FYI: You can rename sheets, just like Excel).  
   C:\Users\Candice\Pictures\new_sheet.png
2. If we really just wanted to get the number of species involved, you can **right-click** on the Dimension: **Species**, and select **“Describe…”**
3. You’ll get a brief overview of the metadata. If you look at **Domain**, it says it’s displaying **20 out of 632 members**. So the quick and dirty answer is 632. But notice that the first item there is Null. *That’s not an animal!* Let’s take a look at what kind of animals involved (and how many null values there are). Close that window and go back to the main screen.
4. Let’s drag the **Number of records** into columns and **Species** into rows. You’ll get bar graphs but it’s hard to tell which species are associated with most collisions.
5. If you click on the sort button, you’ll see what species have the most collisions.  
   
6. You’ll see that the most popular species are:  
   *Unknown bird – medium  
   Unknown bird – small  
   Gulls  
   Unknown bird or bat…*  
   Pretty general eh?
7. A question for you: how many species do you think are *actually* involved?

**iv. How is damage categorized?**

1. One of the many ways to answer this one is to simply drag and drop **Damage** into the Color marks card to make a stacked bar chart:  
     
   This is not the prettiest thing, but you can easily see how many categories are in Damage and you can get a (very) rough idea on how many of the collisions for each species resulted in how much damage.

**v. How many operators are involved in strikes that resulted in aircraft getting destroyed or repair costs going over $2,000,000?**

1. Now we’re looking at the number of records, damage and operator. Let’s drag the **Number of records** into columns and **Operator** into rows (you can sort at this point if you’d like).
2. Now let’s put **Damage** under Color. Since most incidents resulted in no damage or no data was entered (#N/A), the problem is that you can’t see the collisions that resulted in aircraft getting destroyed or repair costs going over $2,000,000. This is where the Filter comes into handy.
3. Drag Damage (the Dimension, not the pill!) into Filters.  
   
4. Tableau will ask you what you’d like to filter out. The ones that are checked will be included, so *un-check everything other than Destroyed and Over $2,000,000*.
5. You’ll probably see something like this now:  
     
   Not a lot of commercial airlines were destroyed – chartered / privately owned aircraft tended to have more incidences of the aircraft getting destroyed by a collision.
6. Bonus question: What animals did those planes run into?

**3. Recap**

Congratulations! You have:

1. Surveyed the data before importing
2. Imported the data into Tableau
3. Learned what the terms are (e.g. pills, shelves, marks card)
4. Answered some basic questions using visual analysis!

If you’re up for some more exploration, move on to **FAA Wildlife Damage to Aircraft Lab 2**!