|  |
| --- |
|  |
| Visualizations based from Wildlife Strikes at Airports |
|  |

|  |
| --- |
| Author: Martin Casey  X-number: X00119025  8-11-2017 |

Contents

[Overview 3](#_Toc497940286)

[Background 3](#_Toc497940287)

[Data sets 3](#_Toc497940288)

[Seven Stages 4](#_Toc497940289)

[Acquire 4](#_Toc497940290)

[Parse 4](#_Toc497940291)

[Filter 6](#_Toc497940292)

[Mine 6](#_Toc497940293)

[Represent 7](#_Toc497940294)

[Total Number of Strikes in % 7](#_Toc497940295)

[Total Days out of Operation 7](#_Toc497940296)

[Airports in the State 8](#_Toc497940297)

[Period of Day 8](#_Toc497940298)

[Animal Categories 9](#_Toc497940299)

[Animals distance from Airport 9](#_Toc497940300)

[Average Cost Per Year 10](#_Toc497940301)

[Airports Per State 10](#_Toc497940302)

[Animals Reported at Airports 11](#_Toc497940303)

[Refine 12](#_Toc497940304)

[Problems & Solutions 13](#_Toc497940305)

[Conclusions 14](#_Toc497940306)

[Recommendations 15](#_Toc497940307)

[References 16](#_Toc497940308)

# *Overview*

The CA in question is designed to take datasets from various sources and create visualisations in Tableau to make aspects of the data more visually noticeable. The purpose of creating visualisations out of the datasets is to make the data more readable to the average user, so that they can easily intercept the data and identify any correlations with the data.

For this CA the author is trying to create visualisations based on wild life strikes at airports, to make it easier to intercept information about the various factors that affect airports, across different regions. The author is hoping to identify the significant disparities between the different animal types that are commonly reported near airlines. The aspects of the wildlife strikes that the author is keeping in mind relate to factors such as number of strikes, animal categories and specific areas where these strikes occur.

# *Background*

The author acquired the dataset on wildlife strikes on airports, after being directed to a link to a tableau dataset site by the author’s lecturer. Initially, the author was working a dataset that related to video games sales over a period of years. However, the author ran into many problems working with the dataset, as there was less and less data available to use, when the dataset was properly put through the seven stages process. Having received advice from the lecturer, the author researched various data sets and interest was piqued regarding how animals could influence airport activities.

Once author found the dataset on the tableau dataset site, they put it through the seven stages process for the CA. The data found in the set was both expansive and organised allowing for many possibilities for the author to create data visualisations.

IT should be noted that as this dataset was found on an official tableau site, its area of interest has existing analysis performed on it throughout the years.

# *Data sets*

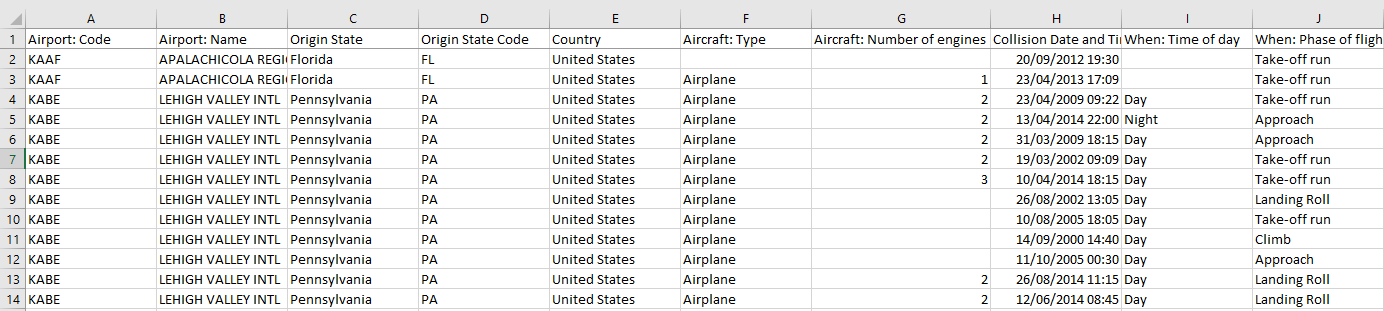
The following is a list of the datasets that the author used during their work on their assignment:

* Dataset: FAA Wildlife Strikes, 2015
* Acquired from: <https://public.tableau.com/en-us/s/resources?qt-overview_resources=1#qt-overview_resources>
* Dataset: Video Game Sales with Ratings
* Acquired from: <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>

# *Seven Stages*

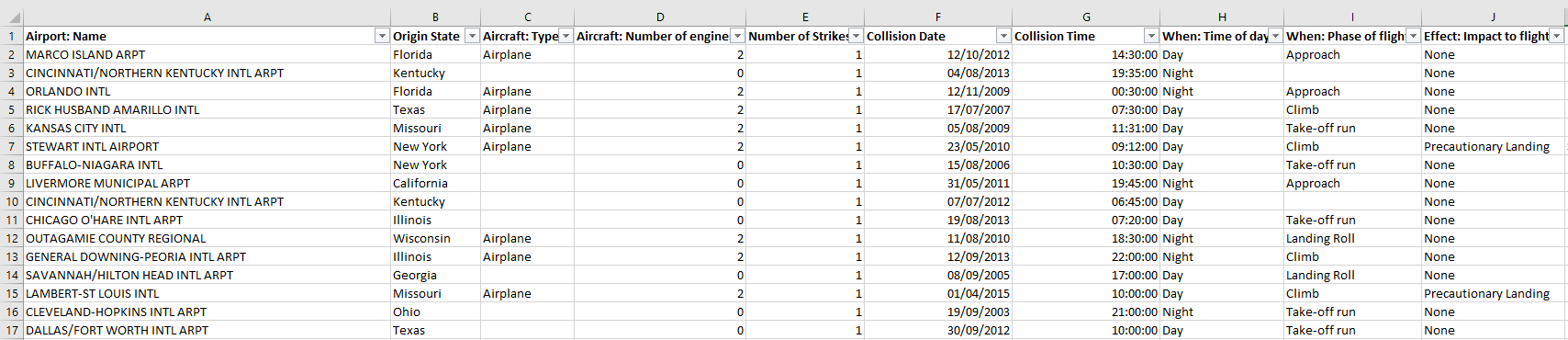
## **Acquire**

For the acquire phase the author acquired a dataset from Tableau based on wildlife strikes at airports. The dataset was suitable for the task at hand and had very few missing values for in some of the columns, which made it easier to work with for the later stages.

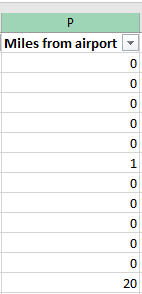


## **Parse**

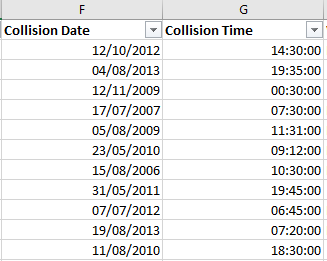
In the parse phase the author began the basics of refining the data for use in tableau. For starters most of the columns were moved around to ensure that the columns that the author felt would be the most relevant would be displayed first. For instance, the columns that detail the Origin State, Aircraft Type, Aircraft No of Engines, Collison Date and Number of Strikes were moved to the forefront as the author felt they would be more relevant for working on visualisations.



Blank cells in all columns were replaced with a 0 if the content consisted of numbers. If the content consisted of letters, the cells were left blank as tableau would make the value of those cells “null” after the import process.



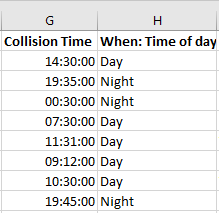
The values in the Collision Date column consisted of a date followed by a time. To allow for more data manipulation options, the author used the features present in excel to split the column into two; one column for the time, and another for the date.



The author noticed that some values in the day column did not line up with the values in the time column. For instance, the day column could read as “Day” when the time column consists of the time “19:30:00”, which would fall under the “Night” iteration of the time of day. To fix this the author created a formula using “if” and “and” statements to register any value in the day column as day if the value in the time column located on the same row was greater than or equal to “6:00:00”, and less than “18:00:00”.

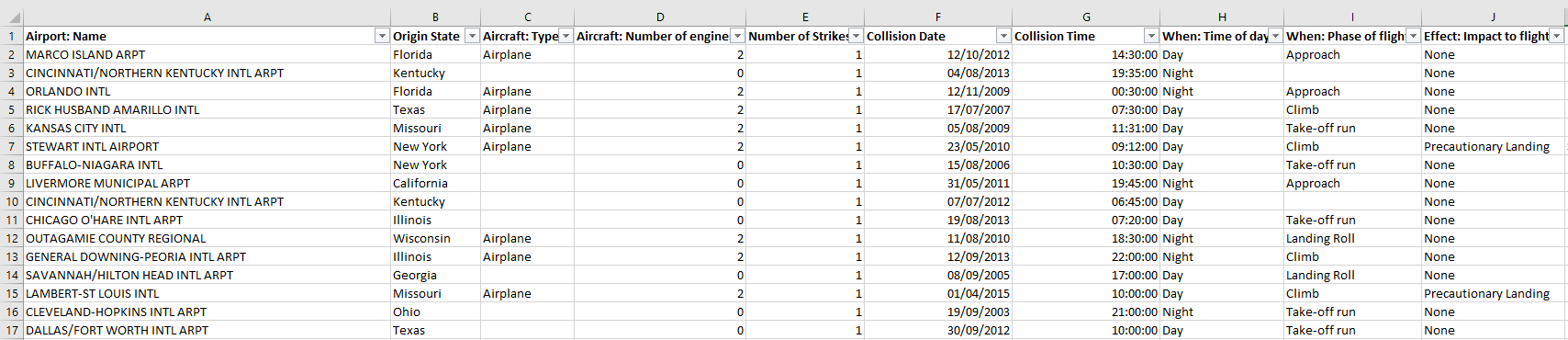


After this formula was applied to the entire day column, all the times of day aligned up correctly with the time values in the time column.



## **Filter**

Everything was left untouched for the filter phase, as based on their past experience with the dataset on video game sales, the author did not want to remove any data as they did not want to severely limit their options for creating visualisations, and run into major issues later on in the assignment.



## **Mine**

The author created two calculated fields for the mine phase of their dataset work. The first field was a field used to calculate the percentage value of strikes across each state. The “Percentage of Total Strike” calculated field allowed the author to show what States consisted of what percentage of the total, for all strikes that happen in airports around the country.



The second calculated field the author created was a field called “Average Cost Per Year”. This field takes the values of the “Total Cost” field and performs a calculation on it to find to find the average cost of airport maintenance. This field is then used with the “Collision Date” field to find the average cost of airport maintenance per year

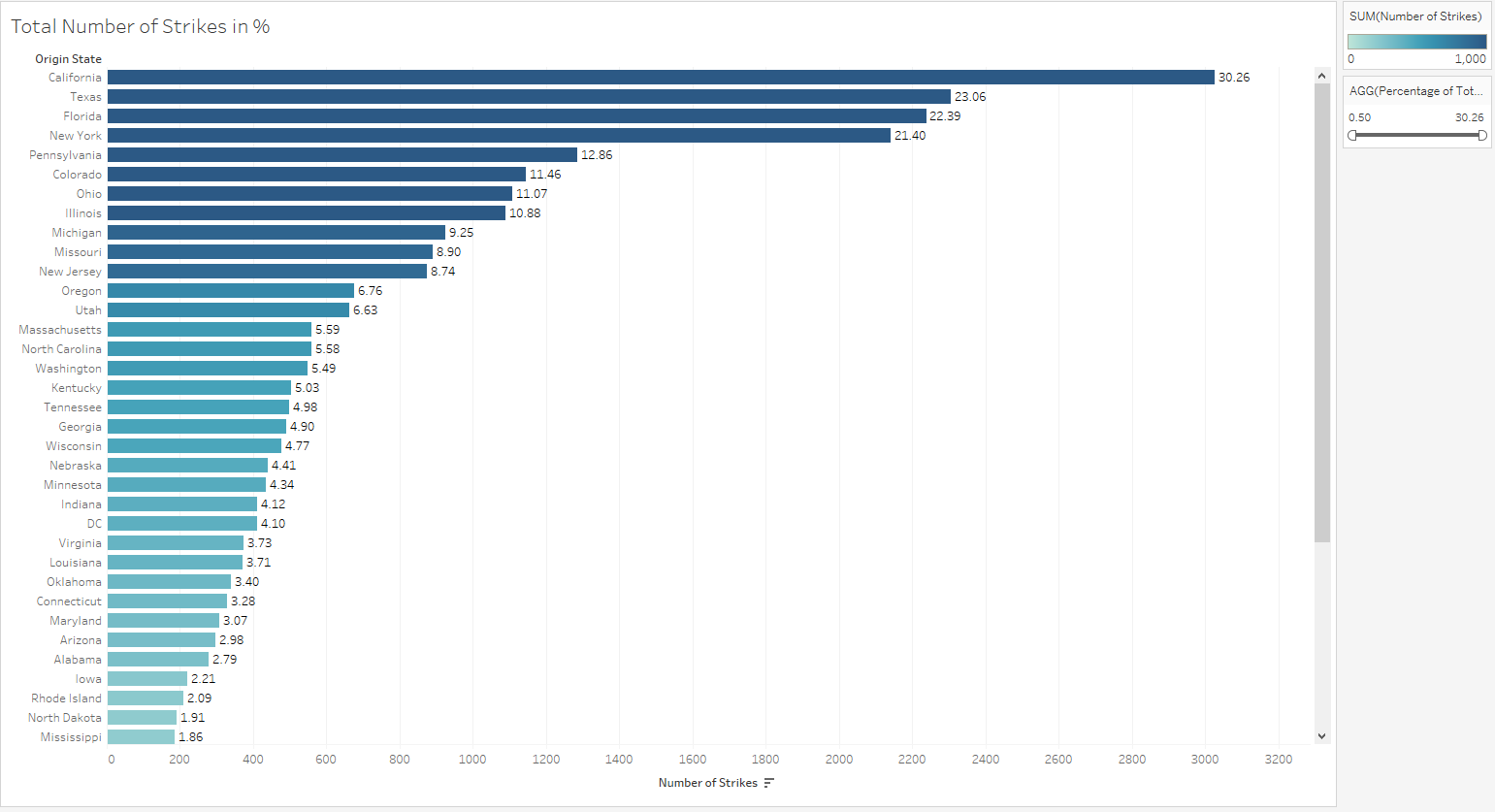


## **Represent**

The author used a total of seven data visualisations for their assessment work. These visualisations were then stored in two dashboards to make them more appeasing to the eye of the user. The details on the visualisations are as follows:

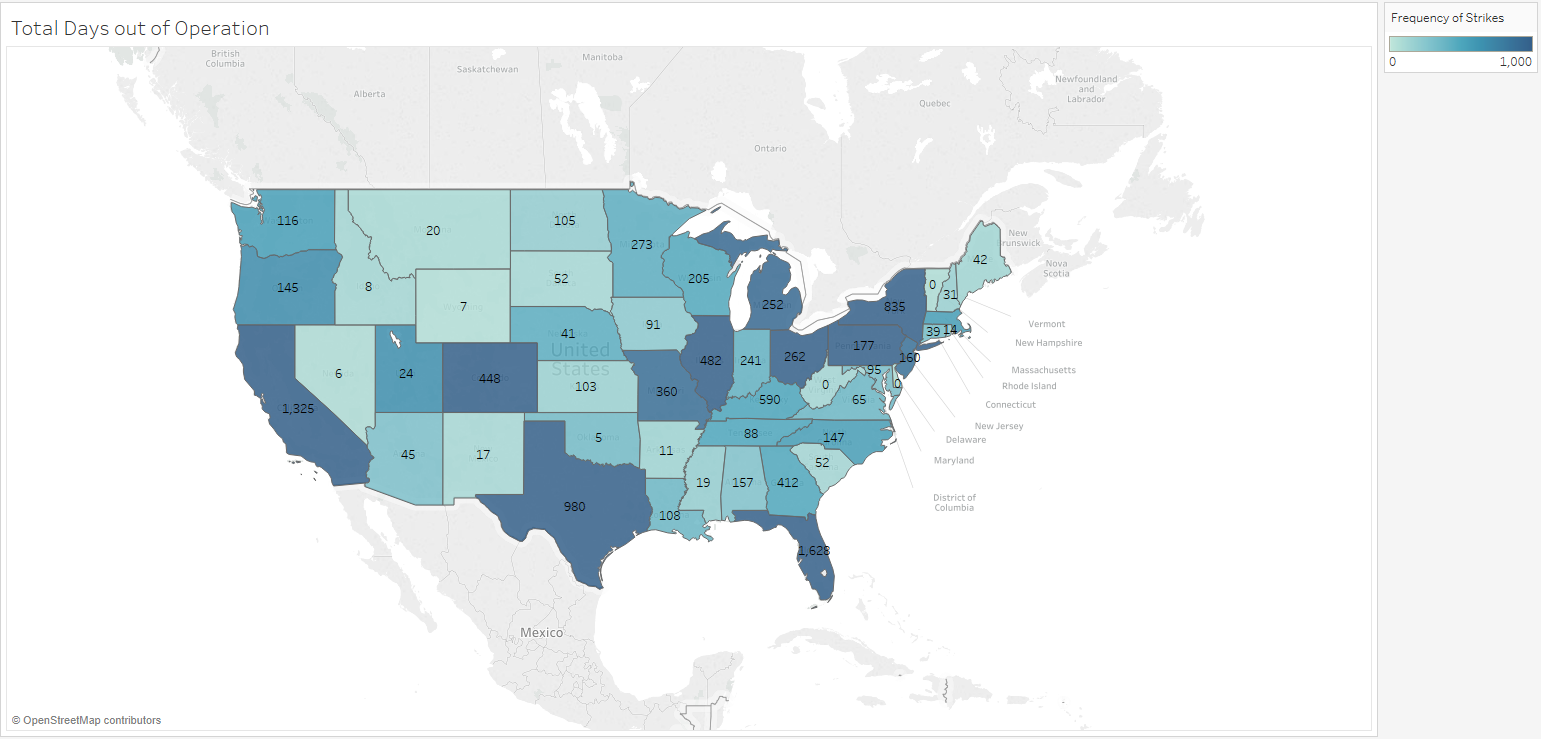
### Total Number of Strikes in %

This visualisation is used to demonstrate to the user the total number of strikes that have occurred at airports per state, and what percentage each state makes up in terms of overall animal strikes. The author decided to display this information using a horizontal bar chart with the states positioned in descending order on the graph based on their number of strikes. The author chose to have the bar chart be displayed horizontally as it allowed for the data to each bar to stretch across the page based on the number of strikes associated with that bar. Also, it was much more visually pleasing, and more manageable to have the name of the states displayed horizontally on the left, as if the bar chart were positioned vertically, the state names would be displayed sideways pointing up from the bottom of the graph, which would not make them easy to read.



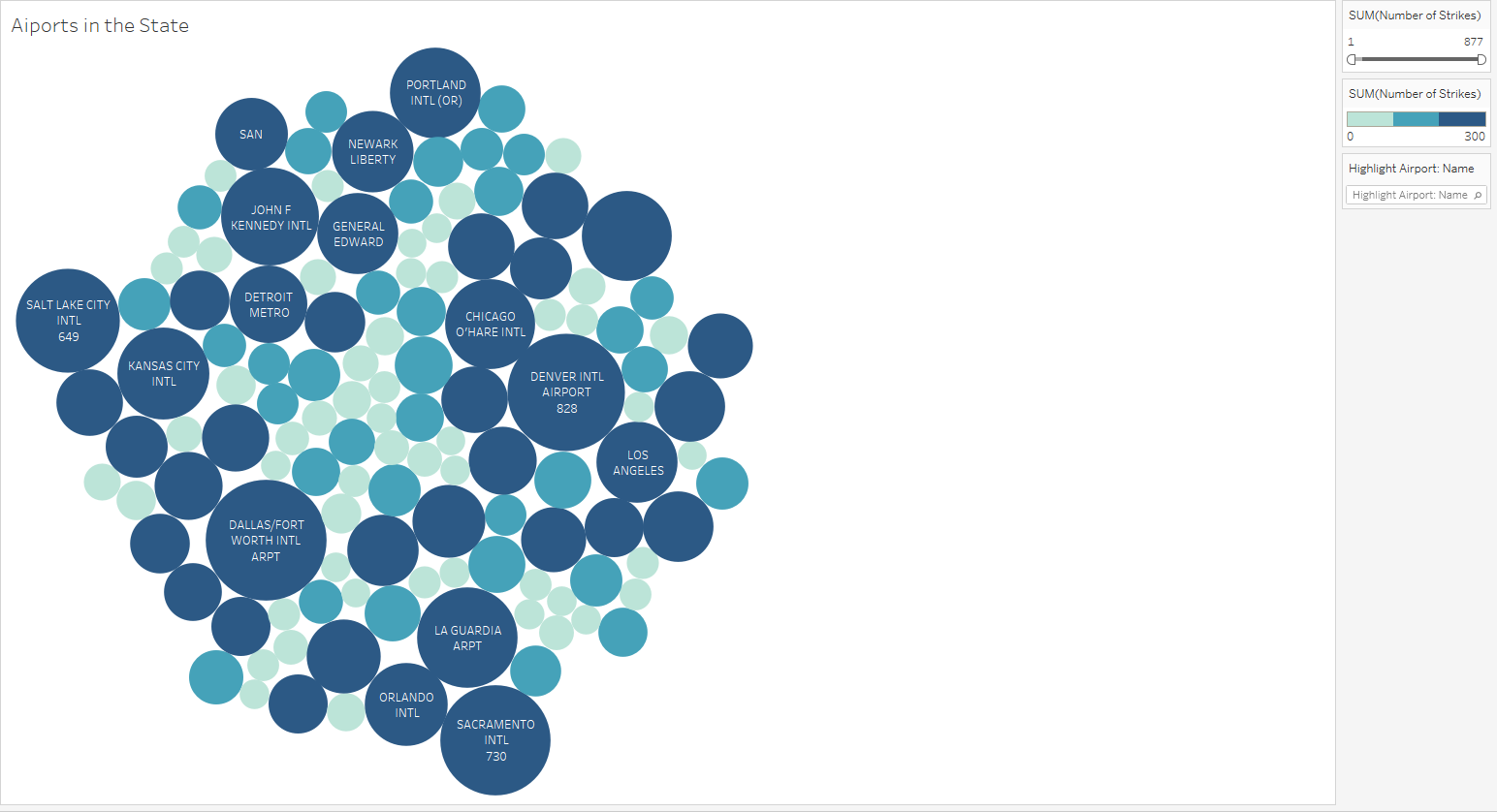
### Total Days out of Operation

This visualisation was used to show the total number of days per state that the airports in those states were unable to provide services. The author felt that it was best to use a map of the country of origin for the states for this visualisation, to show the user at a glance, what states had what number of days of non-active airports. To keep the visualisation in-line with the prior one, the colour and intensity of colour used on each state would be determined by the number of strikes reported by airports per each state. This was to ensure that the visualisations had coherence when used together in a dashboard.



### Airports in the State

This simple bubble graph visualisation is used to show the number of airports in each state, or whole country if view unfiltered. The size and colour of the bubbles is determined by the number of strikes that occur at each airport. So, for instance an airport represented by a large bubble would be an airport that has reported a lot of animal strikes in its time of operation.



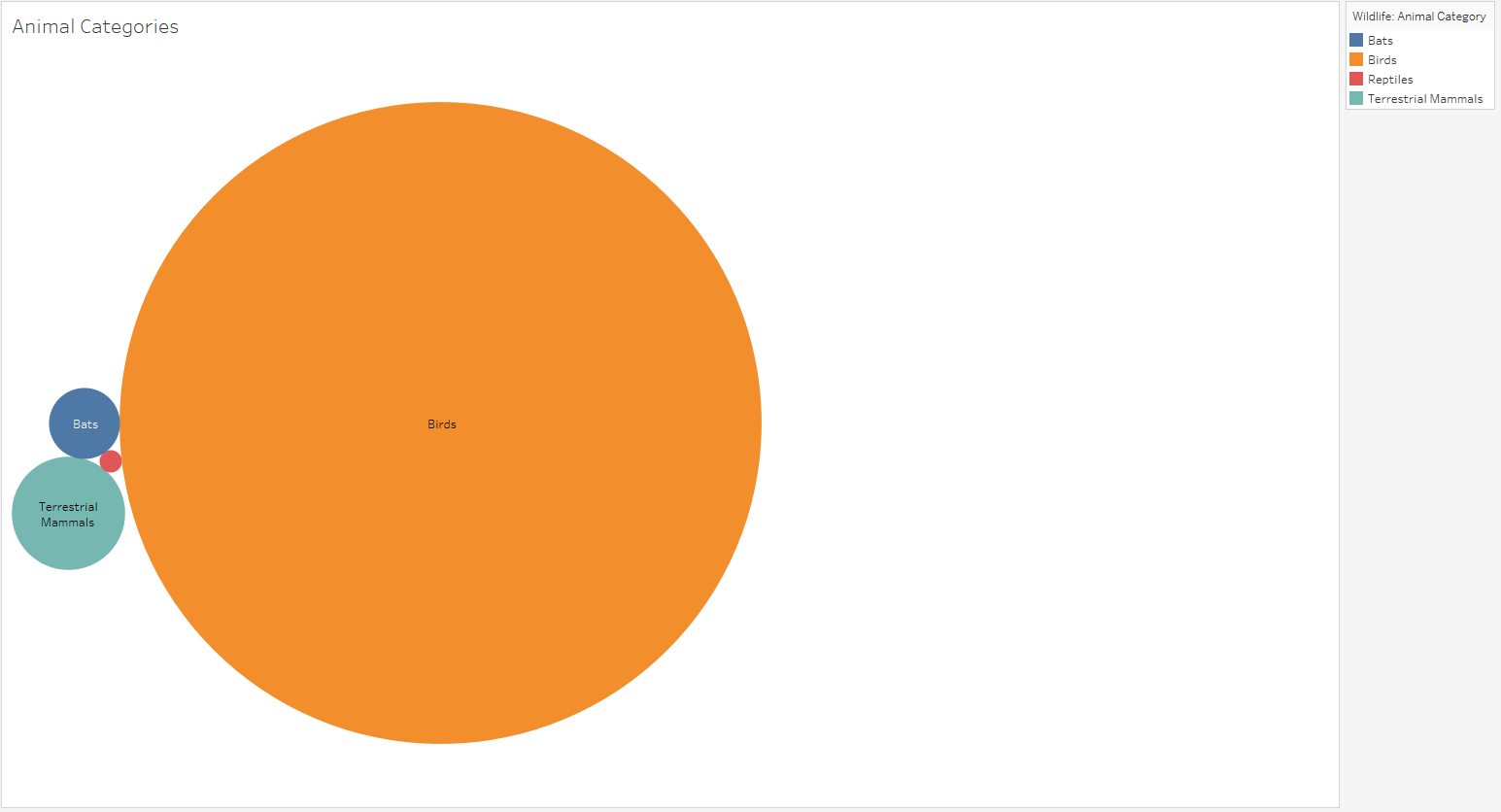
### Period of Day

This text table based visualisation represents the state of day of when animal strikes were reported. There are two values for the state of the day, these are “Day” representing daytime hours form 6:00am to 5:59pm, and “Night” representing nigh time working hours from 6pm to 5:59am. For the animals reported during either how, the table represents them in their respective wildlife species group. This allows the user to differentiate between the groups and in turn examine what species of animal can be found in what group.



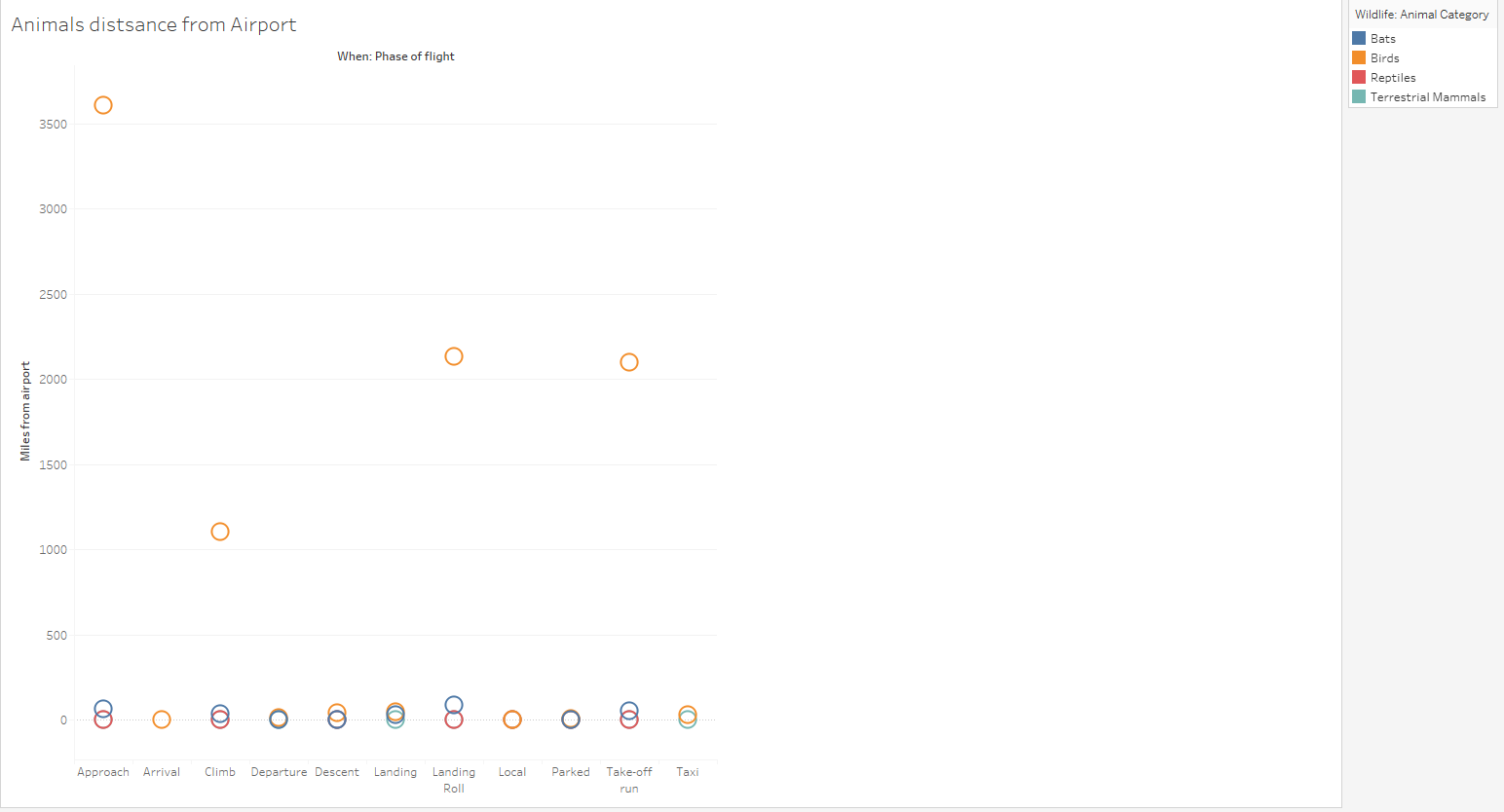
### Animal Categories

The second bubble based visualisation, graph is used to simply display to the user what category of animal is reported in airport wildlife strikes, and which category is more common. Since there is a very small number of categories to show, each bubble in the graph is distinguished by its own unique colour, and its size. A bubbles size is determined based on the number of reports that animals in a category are reported on. The graph is kept simple as it is mainly used as a trigger in the dashboard it is contained within.



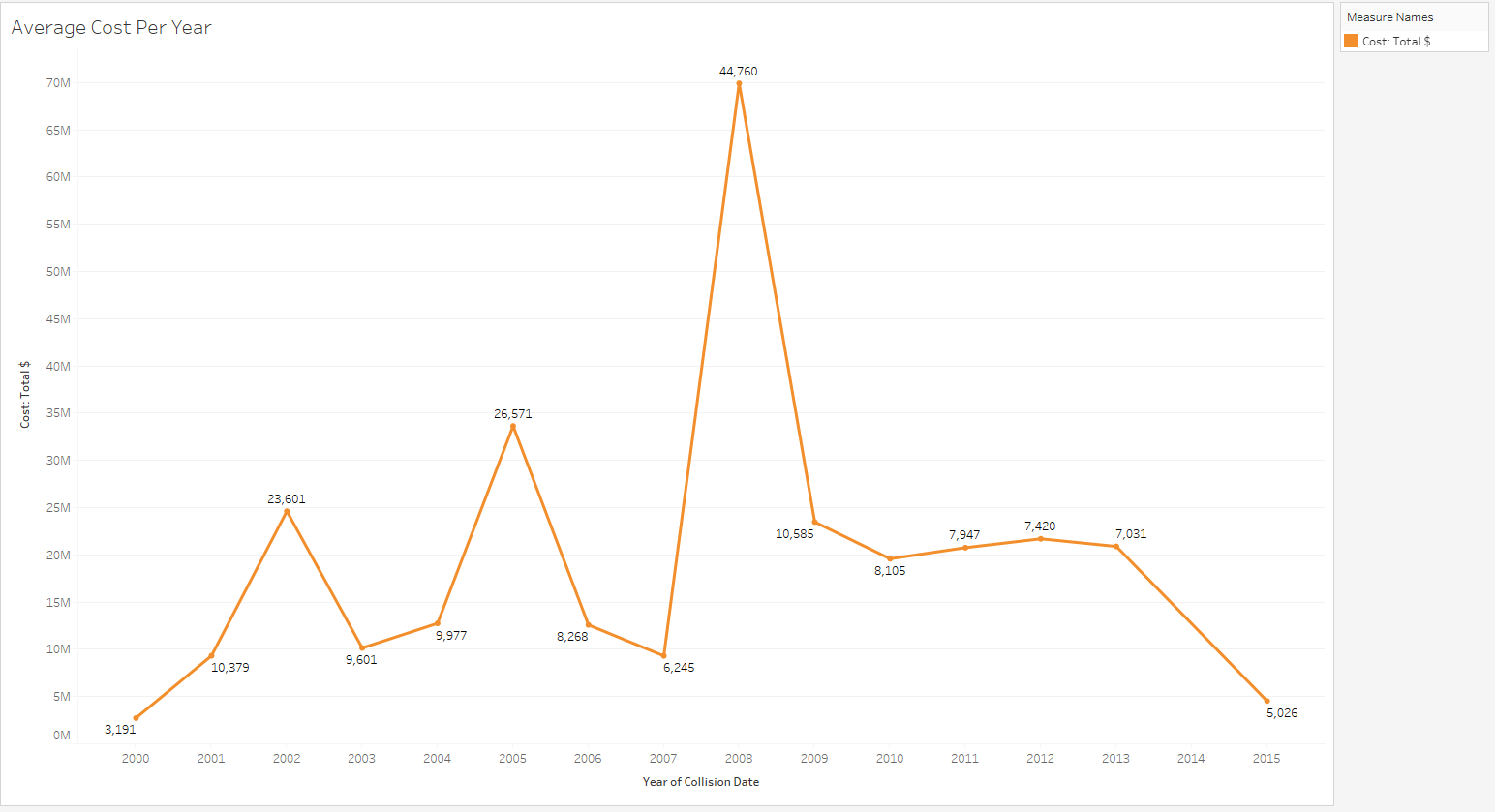
### Animals distance from Airport

A circle view was the graph of choice for this visualisation. This graph is used to display to the user what the distance each category of animal was from the airport when the strike reports were made, and what phase of flight the planes in the airport were operating in when the reports were made. Each category of animal is represented by a coloured circle, which used the same colours for each category as the prior graph. The author decided to use circles to represent this data as the user can then simply focus directly on the relevant information.



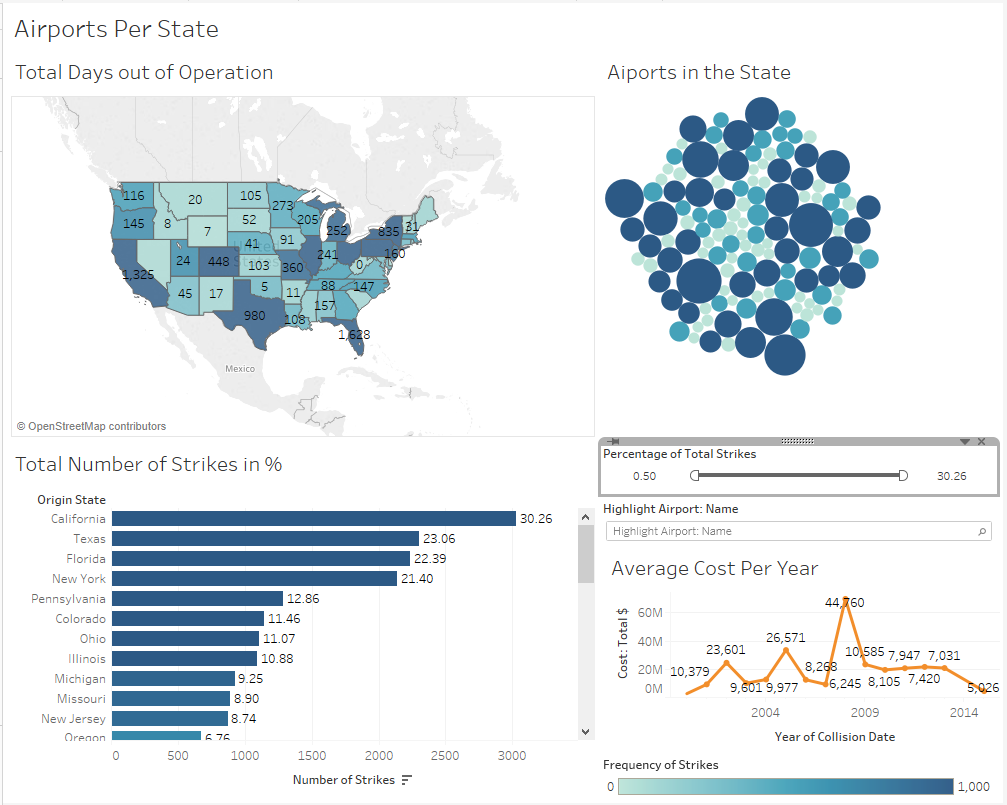
### Average Cost Per Year

This simple line graph visualisation shows the user the rise and fall of the average cost per year of maintenance of an airport in the states, or country if view unfiltered. The author decided to stick with the default orange colour that Tableau provided, as the colour of the line easily stood out on the graph. The author created a calculated field called “Average Cost Per Year” to display to the user what the average cost of airport maintenance for a given year was along with the total cost, for easy comparison. Rather than have the average cost be displayed as another line on the graph, the author found it more suitable and simpler, to display the average cost at each year point in the graph using the label feature in Tableau.



### Airports Per State

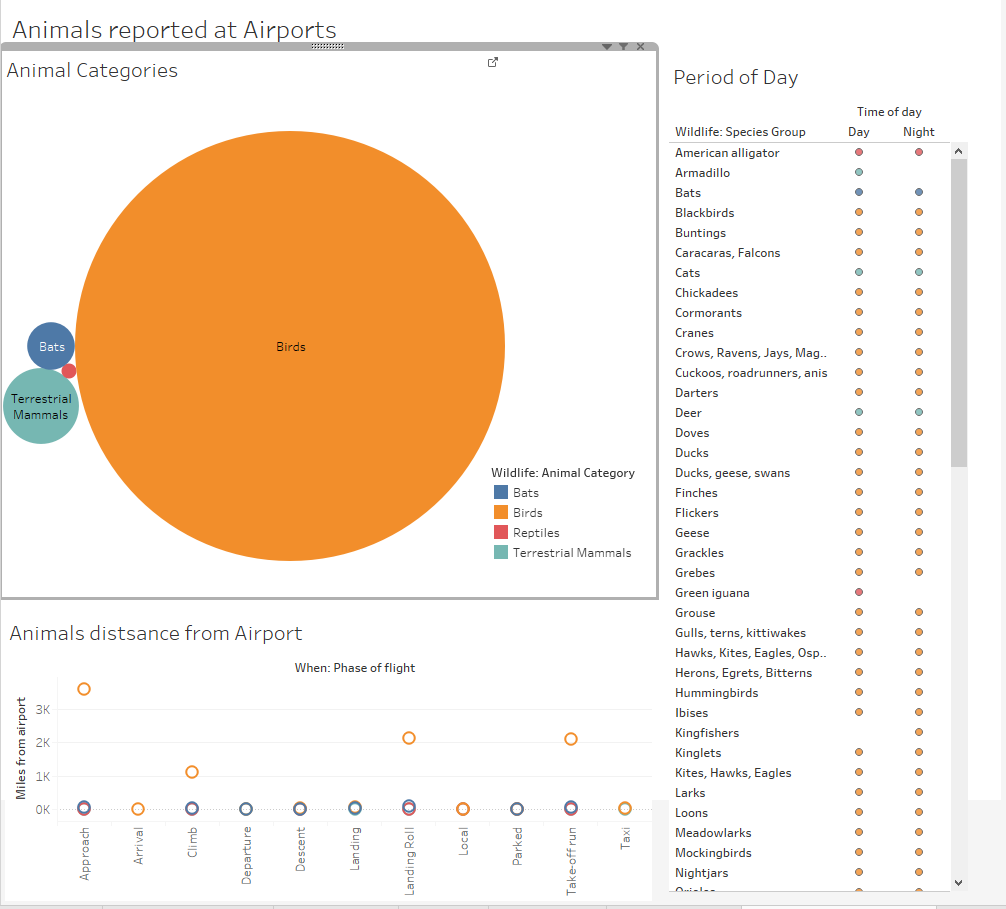
This dash board consists of the visualisations, “Total Days out of Operation”, “Total Number of Strikes in %”, “Airports in State” and “Average Cost Per Year”. The dash board has been modified to use features to both filter and highlight aspects of the other visualisations, based on the state that the user selects. For instance, if the user selects the state “California” the “California” column in the “Total Number of Strikes in %” visualisation on the dash board will be highlighted. At the same time the “Airports in State” visualisation along with “Average Cost Per Year”, will be filtered to only show data on airports that are in “California”.



To give more options to the users filter toggle for the “Percentage of Total Strikes” not only affects cells in the “Total Number of Strikes in %” visualisation by removing/inserting them to the users view based on strike numbers, but if will also remove/insert the values of the other visualisations as well, only showing the range of values that the user wants. Below this toggle filter is another filter, a text box which the user can use to highlight a airport in the “Airports in State” visualisation, by filtering out any airports that do not match the name inputted by the user. The filter does also try to show the user names of the airports while they are typing, just for more accessibility. Finally, below the “Average Cost Per Year” visualisation is a colour bar, which shows the what the range of colours on the first two graphs are and the volume of strikes each colour represents.

### Animals Reported at Airports

This dash board consists of the visualisations “Animal Categories”, “State of Day” and “Animals distance from Airport”. The “Animal Categories” visualisation acts of the main point of interaction for the dashboard, as the actions the dashboard uses are triggered when the users interacts with this visualisation. For instance, should the user click on the bubble labelled “Birds” in the visualisation. The state of day visualisation would be filtered to only show the instances of different groups of birds that where reported on and what state of day the report was made. While the “Animals distance from Airport” visualisation would only show the distances of birds from the airport and what states the airplanes were in when the reports were made.



Since the colour scheme for each animal category remains consistent across each of the three visualisations, the author felt that it was only needed for the legend for the categories, which detail what each colour represents, to be displayed in a clear spot on the bubble graph, close to the centre of the screen. So that it is easy for the uses eye to catch.

## **Refine**

For finishing up the work on the visualisations, the author spent a small amount of time refining the different aspects of the visualisations. The main aspect that the author needed to refine when it came to the visualisations was their colours. The author found that some of the visualisations became straining on the eyes of the user over a prolonged period, due to the colours they used and the intensity of those colours. The author also found that most of the colours weren’t uniform with each other when the visualisations were combined onto dashboards. To fix this issue the author considered the various colour options in tableau, and assigned colour plates to the visualisations that were less strenuous on the user’s eyes. Also, they made sure that the colours in the visualisations were consistent with each other when used on a dashboard.

# *Problems & Solutions*

Over the course of working on the assignment the author ran into a few problems, both large and small, while trying to create visualisations. These are listed as follows

**Problem:** Original dataset became unsuitable

Initially the author was working on the assessment using a dataset that related to video game sales. The dataset focused on the sales of video games across many years, detailing different factors related to the games such as their publishers, developers, platforms, age rating, and sales across different regions. However, after the dataset was put through the seven stages process the author realised that a most of the data was removed from the dataset, most of the cells in the dataset had invalid values, such as too many blank cells, making it difficult to form a coherent assumption on what values blank cells should contain. This made it difficult to make any visualisations of the dataset without constantly re-using the same values repeatedly.

**Solution:** Acquire new dataset

After the author informed their lecturer of the issue they were having with their dataset, the lecture directed the author to a set of links on the Interactive Media Design Moodle page. These links pointed the author to two possible sources to acquire information. After the author searched through the datasets on the tableau page, they found a dataset on wildlife strikes at airports, and found that to be the most intriguing of datasets the author had come across. The download the dataset and began working on readying it for visualisation shortly after.

**Problem:** Separating time from date

During the parse phase on the wildlife strikes dataset, the author noticed that collision times and date of collision times were all located in the same column. The author felt it would be more effect or the date and times to be in separate columns, as that would allow for more visualisation options down the road. However, the author was unsure about how to separate the values, without going through each row individually, which would be too time consuming.

**Solution:** Create function to perform separation task

The author looked online for how formulas in Microsoft excel worked, so that they could achieve their task. They found some general guidelines, but were still confused on the topic, so they emailed their lecture to ask for advice. Their lecture responded back to them, providing them with an email link to a site which explained how formulas and functions worked in excel. The author looked up how the functions worked and tried to create a simple one on a test sheet to first see how the functions worked in person.

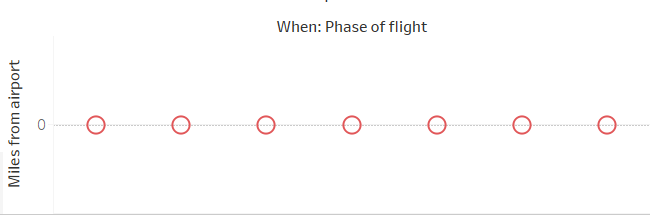
However, the author could not get the functions for a while, until they came across a solution online which explained to them how they show structure their formulas in excel based on the region they are operating in. After the author implemented this advice they got the function working, and then looked up online how to include an “AND” statement in their “IF” function. Once they found out this information they could separate the data into two separate columns.

# *Conclusions*

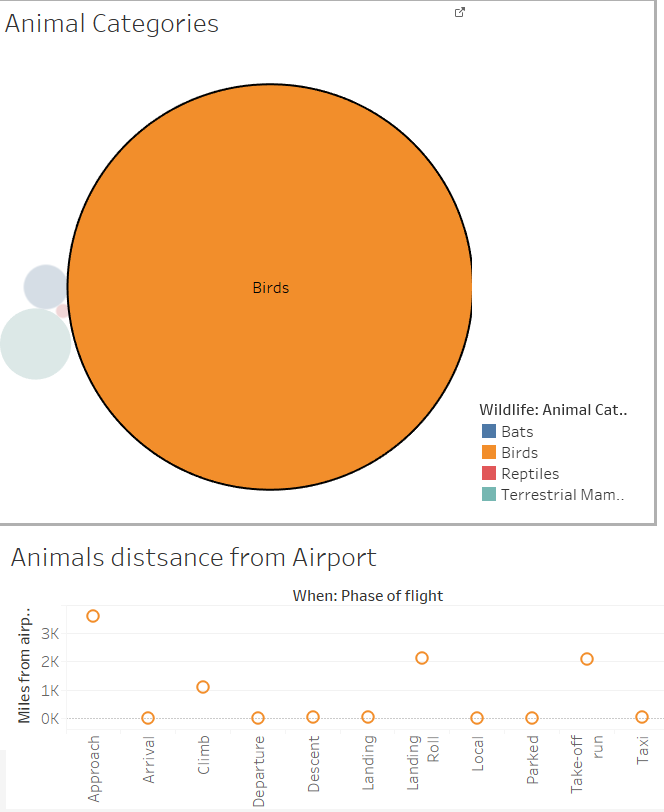
After many trials and tribulations, the author managed to complete their work on the visualisations based from the wildlife strikes at airports in the United States of America. Prior to working on these visualisations that author only had a very basic understanding as to what animals could affect aircrafts in airports. After creating and analysing the visualisations the author found out that it was more than just birds that affected the aircraft in airports during their various phases of flights. For instance, the author was surprised to discover that animals that fall under the reptilian category appeared commonly enough in airports, that they be counted in the animal category as an animal type that disrupts the flight of aircrafts.

The author found this surprising as when viewing the bubble graph visualisation which shows the quantity of strikes reported for the different animal categories, the reptilian category made up the smallest bubble in the entire graph. Yet specifies from this category proved to be consistent enough in appearing at airports at all phases of flight for an aircraft, that they could still make a notable impact on the flight.





It should also be mentioned that the visualisations also confirmed data to the author which they expected. Such as the “Animal Categories” visualisation which unsurprisingly confirmed to the author that birds where the most common category of animal reported to affect airlines. As well as being the primary category of animal to be reported on when they are a long distance away from the airport. Which makes sense of course as they are also the category of animal that spends long periods of time in the sky, which is where the aircrafts need go to progress with their journeys.



Another point that the author found interesting in their discoveries from the data visualisations was that was that out of all the States in the United States of America, four of them amassed to take up more than seventy percent of the total wildlife strikes that are reported.



## **Recommendations**

When it comes to further analysis in this field of data, the author would recommend researching into the possible number of populations that the various airports house on a day to day basis. Research into this area could be used to find out if factors such as population could affect areas such as the cost of maintenance for an airport. As well as what the average amount of downtime days would be for airports.

The author would also recommend researching into data on geographic regions and behavioural patterns for the various animals found to affect airports. This could be used to find out whether there is any correlation activities of the various animals and the regions near an airport, to understand if say mating season for certain mammals in an area would make them more likely to appear at an airport located near a region.

# *References*

* Tableau Public (2017). Resources/How-to-Videos [online]. Available at: <https://public.tableau.com/en-us/s/resources?qt-overview_resources=1#qt-overview_resources> [Accessed 2nd November 2017]]
* Kaggle.com (2017). Featured Dataset Video Game Sales with Ratings[online]. Available at: <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings> [Accessed 8th November 2017]]
* Excel Forum (2017). “We found a problem with this formula” Error? [online]. Available at: <https://www.excelforum.com/excel-formulas-and-functions/1011332-we-found-a-problem-with-this-formula-error.html> [Accessed 8th November 2017]
* Exceljet (2017). If this AND that [online]. Available at: <https://exceljet.net/formula/if-this-and-that> [Accessed 8th November 2017]