



# MONASH University

## FIT5147 Data Exploration and Visualisation

Semester 1, 2023

### Programming Exercise 1: Tableau Public

Name	Nazmus Sakib
Student ID	33361991

## Contents

Overview .....	3
Error Handling .....	3
• Outliers.....	3
• Duplicate Data.....	4
• Inconsistent Data .....	5
Question 1.....	5
Answer .....	5
Question 2.....	7
Answer .....	7

The dataset that had been given to us needs to be upload/imported in Tableau. It has a total of 6995 rows and 12 columns of data. When it is imported in the Tableau Desktop it looks like the image in the figure below (Figure 1).

PEI_frog_data.csv		12 fields 5995 rows										100 rows	
Name		#	#										
PEI_frog_data.csv		PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv	PEI_frog_data.csv
		F1	Unique ID	Date	Time Start	Latitude	Longitude	Type of observation	Common name	Scientific name	Family	Genus	Species
		336	10097	26-Mar-01 12:00:00 AM	30-Dec-99 7:40:00 AM	-3766500	144.832000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		337	10096	26-Mar-01 12:00:00 AM	30-Dec-99 7:40:00 AM	-3765700	144.839000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		409	10073	11-Jun-01 12:00:00 AM	30-Dec-99 8:00:00 AM	-37680200	144.839000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		410	10072	11-Jun-01 12:00:00 AM	30-Dec-99 8:00:00 AM	-37662200	144.832000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		411	10070	11-Jun-01 12:00:00 AM	30-Dec-99 8:00:00 AM	-37659400	144.830000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		412	10070	11-Jun-01 12:00:00 AM	30-Dec-99 8:00:00 AM	-37652700	144.840000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		419	10081	18-Jun-01 12:00:00 AM	30-Dec-99 8:00:00 AM	-37688900	144.843000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		420	10080	18-Jun-01 12:00:00 AM	30-Dec-99 8:00:00 AM	-37663200	144.829000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		425	10086	25-Jun-01 12:00:00 AM	30-Dec-99 7:50:00 AM	-37680200	144.839000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		1986	11748	07-Apr-03 12:00:00 AM	30-Dec-99 9:30:00 AM	-37652700	144.840000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		2001	11773	02-Jun-03 12:00:00 AM	30-Dec-99 8:00:00 AM	-37652700	144.840000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		2008	11780	07-Jul-03 12:00:00 AM	30-Dec-99 8:00:00 AM	-37652700	144.840000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		2026	11788	04-Aug-03 12:00:00 AM	30-Dec-99 8:10:00 AM	-37680200	144.839000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		2398	12612	06-Oct-03 12:00:00 AM	30-Dec-99 8:00:00 AM	-37652700	144.840000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		2484	12610	05-Apr-04 12:00:00 AM	30-Dec-99 8:00:00 AM	-37652700	144.840000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera
		2885	12650	03-May-04 12:00:00 AM	30-Dec-99 8:05:00 AM	-37682300	144.837000	Frog census (AIW code)	Eastern Common Froglet	Crinia signifera	Myobatrachidae	Crinia	signifera

Figure 1: Overview of the raw PE1\_frog\_data.csv

## Error Handling

After the data is analyzed it can be observed that there are few kinds of errors in the raw data. The type of errors are described as below.

- **Outliers:** There are few outliers in the dataset. For example most of the longitude values are positive which represents the eastern hemisphere, but there is a value with negative longitude which actually represents the western hemisphere. Same goes for latitude, there is a positive latitude which represents the northern hemisphere whereas all the other values are negative latitude. There are also outliers in the Date where the Date was from 1916 which is wrong because the data was from the year 2000. The figure below (Figure 2) shows the outlier in Longitude. The outliers are handled but not considering them in the data that is made ready to analyze. There the outliers are omitted in the cleaning process.

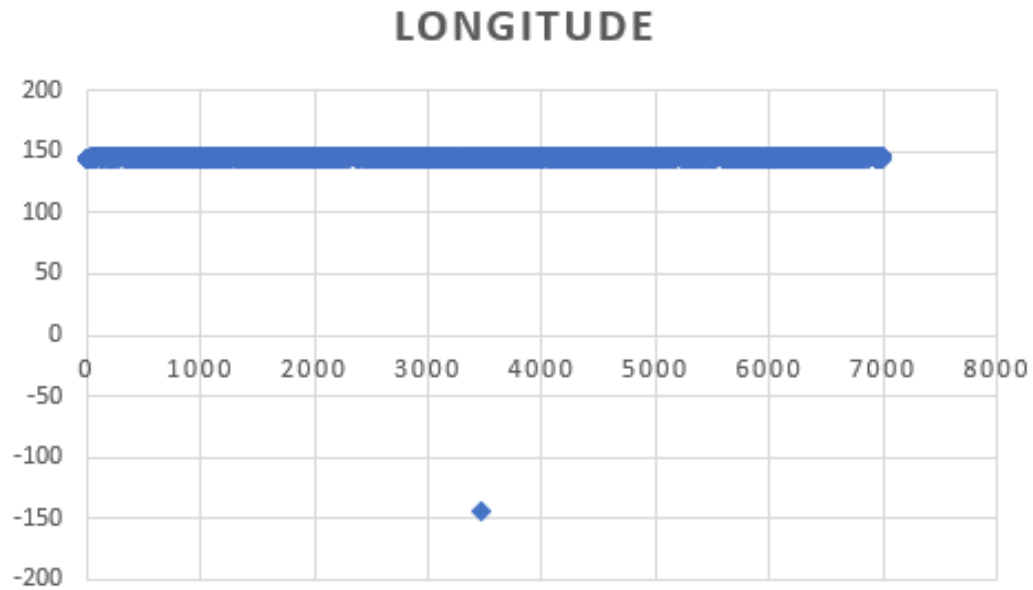


Figure 2: Outlier in Longitude

- **Duplicate Data:** In Common\_name column there are values which are same but have been put with different common names, so when analyzing the value will be stored as different values. For example the figure below (Figure 3) shows the values which are repeated. The Red Box shows the same value but with lower case letters and the Yellow Box shows the same value but mistakenly with a sign in the name. These errors are handled by turning them into the same Common\_name attribute as they are not different values.

Sheet 1

Common name	November 1916 1899		March 2001 1899		June 2001 1899		April ; 18:
Common Spadefoot Toad							
Eastern Banjo Frog							
Eastern Common Froglet							
Eastern Dwarf Tree Frog							
Eastern Sign-bearing Froglet							
Growling Grass Frog							
Growling grass frog							
Peron's Tree Frog							
Peron&€™s Tree Frog							
Red-groined Froglet							
Southern Brown Tree Frog							
Southern Toadlet							
Spotted Marsh Frog							
Striped Marsh Frog							
Victorian Smooth Froglet							
Whistling Tree Frog							
	0	0 10 20	0	0 10 20	0	0 10 20	0
	Hour of Date	Hour of Time St..	Hour of Date	Hour of Time St..	Hour of Date	Hour of Time St..	Hour of Date

Figure 3: Repetition of Common\_name Attribute

- **Inconsistent Data:** The two columns' Date and Time\_Start are inconsistent with each other. It is observed that the Date is in the Date Time format but the time for all the dates are set to 00:00:00+00 for all records and the year on Time\_Start is set to 1899/12/30 for all records, which does not make sense as the observations are made from the year 2000. So these errors are handled by turning Date column into only dates instead of DateTime and turning the column of Time\_Start to only time instead of DateTime. Then they are merged together to get the appropriate DateTime Format of the observations made. The figure below (Figure 4) shows the errors seen in the dataset.

Sheet 1

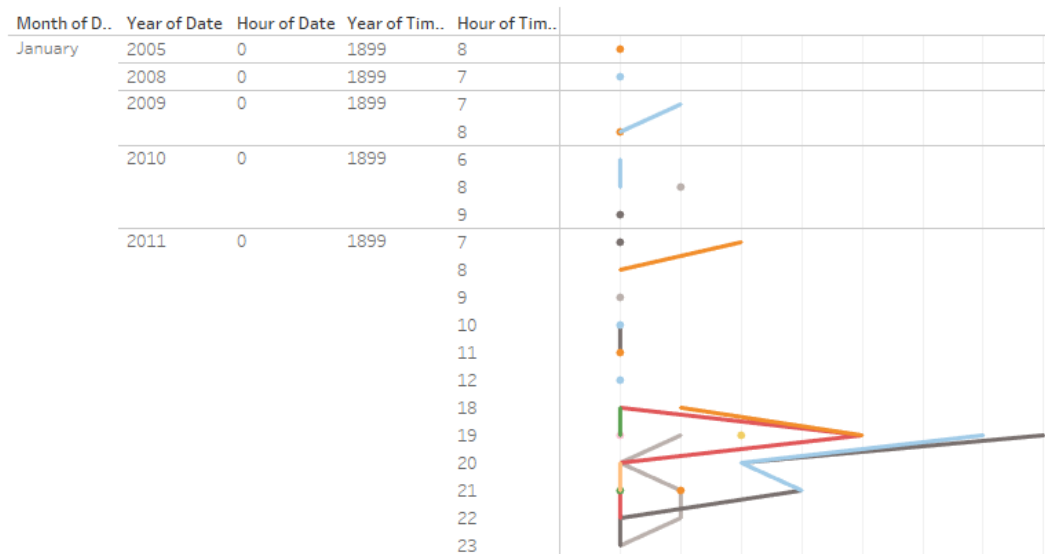


Figure 4: DateTime Format Error for Date and Time\_Start Column

## Question 1

Compare and contrast when different types of frogs were observed. Consider this on both an hourly and a monthly timescale. What does this suggest about the behavior of the researchers or the frogs?

**Answer:** When the data of the frogs are visualized in the graph on both hourly and monthly timescales, on hourly graph it can be seen that at midnight 12 am the dumerilli species are the lowest seen and signifera species can be seen the most but very less in numbers, only 7. The frogs start to show up mostly at 4 am, signifera in 28, peronii in 20, demerilli in 13 and least is haswelli in 1 quantity, and the quantity decreases gradually for the next 2 hours for all the species. Then after 6 the quantity rises very steeply. At 8am signifera are seen the most at 519, peronii in second at 66, then dumerilli at 48, and tasmaniensis at 46, the others are as low as it gets. But with in the next 2 hours signifera drops down to 89 only and the others also drop gradually but not like signifera. All the frogs keeps on dropping until 3 pm, then at 5 the spike is seen again but in lower numbers than before, again signifera at 236, ewiingii at 216, dumerilli at 146, peronii at 122, verreauxii at 91 and the others are very low at quantity. But after

8pm all the frogs keep on decreasing until the end of the night. The figure below (Figure 5) shows the Count of Species against Hour Start.

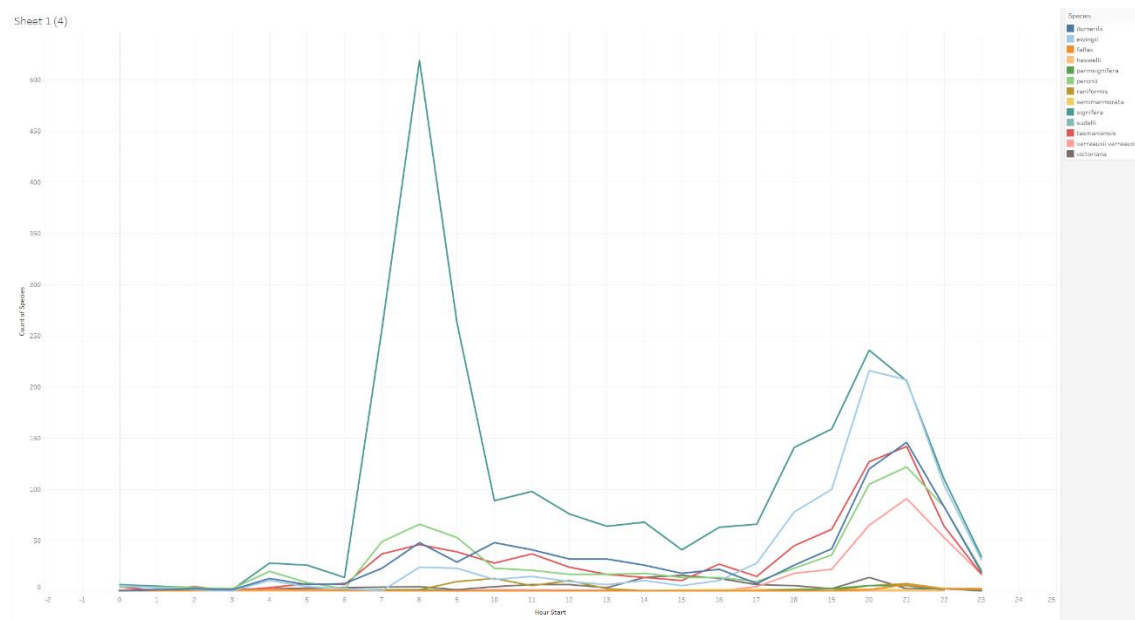


Figure 5: Count of Species against Hour Start

Then again for the monthly basis chart we can see that Victoriana, haswelli and fallax are very rare type of frogs which are seen in very few number all over the year, the most common is signifera which are most seen in April, August and October. Dumerilli is mostly seen in October and November, peronii is also seen in October and November. Tasmaniensis and Ewingii are seen in October only and the others like raniformis and parinsignifera are only seen in October and November and by December all of them goes back to shelter. The figure below (Figure 6) demonstrates the data.

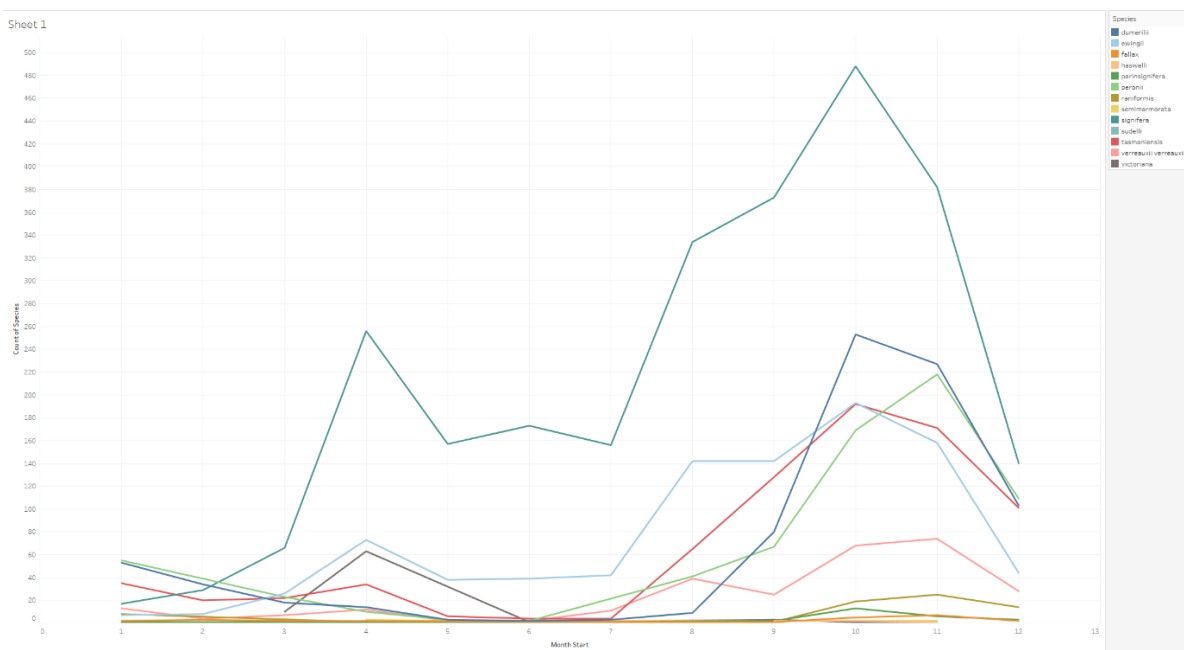


Figure 6: Count of Species against Month Start

## Question 2

Compare and contrast the “Type of Observation” used to gather the data. How does this variable support, challenge or change your conclusions to the first question?

**Answer:** The type of observations matter but it is consistent with the species seen. For example the most frogs are seen at the end of the year with at October to November by Opportunistic Recording at 630, Frog Census at 389, Frog Census (AVW) at 380, Heard during bird survey at 151. The most difficult type of observations is fixed point count, incidental and Sound Recording. By these methods frogs are rarely seen. Why these graphs are chosen because the most of the attributes are constant and to display the trend line graph can be the best visualization tool to understand the differences between months and hourly captures. The other types of visualizations does not give us the full picture from the graph but from the line graph with eh legends shown at the side the viewer can see the whole picture of the analysis at one glance. The figure below (Figure 7) shows the graph of Count of Type of Observation against Month Start.

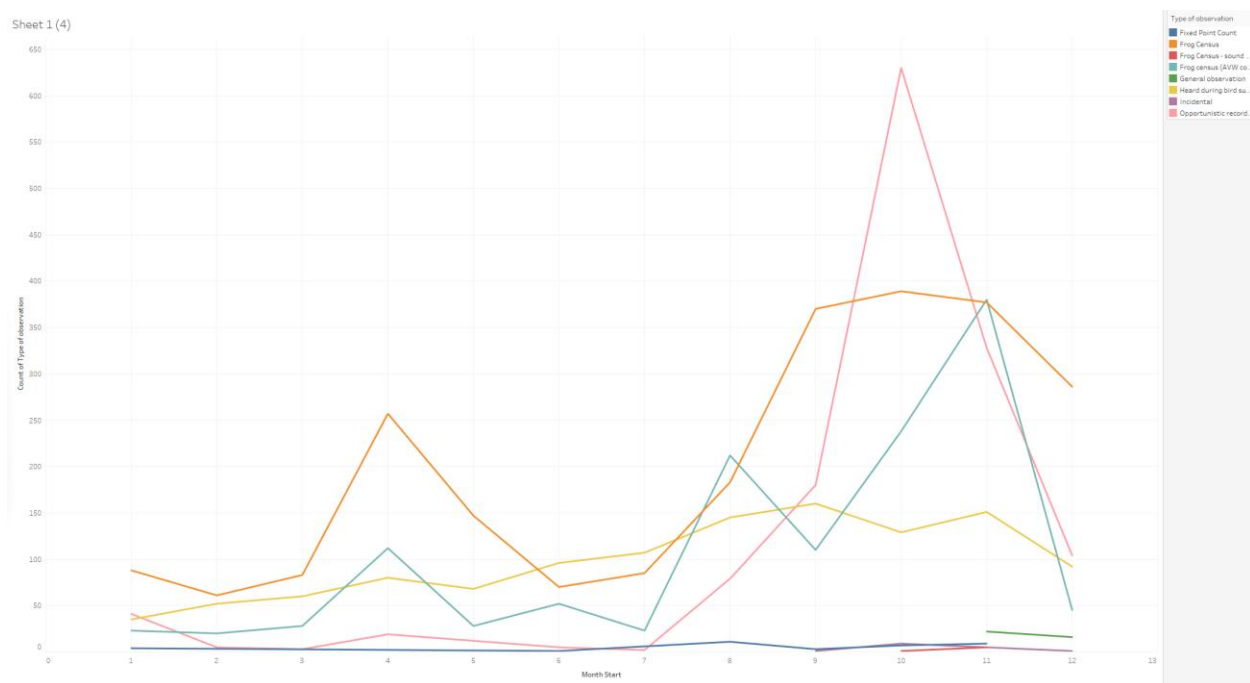


Figure 7: Count of Type of Observation against Month Start