

FIT5147 Data Exploration and Visualisation

Semester 1, 2023

Programming exercise 1: Tableau Public

By: Pragy Parashar

Student ID: 31940757

Tutor Name: Deep Mendha

Data Analysis:

Name	Field Name	Physical Table	Remot...
PE1_frog_data.csv	F1	PE1_frog_data.csv	F1
Unique ID	Unique ID	PE1_frog_data.csv	Unique...
Date	Date	PE1_frog_data.csv	Date
Time Start	Time Start	PE1_frog_data.csv	Time_St...
Latitude	Latitude	PE1_frog_data.csv	Latitude
Longitude	Longitude	PE1_frog_data.csv	Longitude
Type of observation	Type of observation	PE1_frog_data.csv	Type_of...
Common name	Common name	PE1_frog_data.csv	Commo...
Scientific name	Scientific name	PE1_frog_data.csv	Scientifi...
Family	Family	PE1_frog_data.csv	Family
Genus	Genus	PE1_frog_data.csv	Genus

Unique ID	Date	Time Start	Latitude	Longitude
336	2001/03/26 00:00:00+00	1899/12/30 07:40:00+00	-37666500	144.8320
337	2001/03/26 00:00:00+00	1899/12/30 07:40:00+00	-37657100	144.8190
409	2001/06/11 00:00:00+00	1899/12/30 08:00:00+00	-37680200	144.8390
410	2001/06/11 00:00:00+00	1899/12/30 08:00:00+00	-37662200	144.8220
411	2001/06/11 00:00:00+00	1899/12/30 08:00:00+00	-37659400	144.8310
412	2001/06/11 00:00:00+00	1899/12/30 08:00:00+00	-37652700	144.8410
419	2001/06/18 00:00:00+00	1899/12/30 08:00:00+00	-37668900	144.8430
420	2001/06/18 00:00:00+00	1899/12/30 08:00:00+00	-37663200	144.8290
425	2001/06/25 00:00:00+00	1899/12/30 07:50:00+00	-37680200	144.8390
1986	2003/04/07 00:00:00+00	1899/12/30 09:30:00+00	-37652700	144.8410
2011	2003/06/02 00:00:00+00	1899/12/30 08:00:00+00	-37652700	144.8410
2018	2003/07/07 00:00:00+00	1899/12/30 08:00:00+00	-37652700	144.8410
2026	2003/08/04 00:00:00+00	1899/12/30 08:10:00+00	-37680200	144.8390
2398	2003/10/06 00:00:00+00	1899/12/30 08:00:00+00	-37652700	144.8410
2849	2004/04/05 00:00:00+00	1899/12/30 08:00:00+00	-37652700	144.8410

Fig 1: Dataset loaded in Tableau.

The PE1 Frog Dataset is loaded in Tableau Public as a csv file. Initial data analysis is performed on the data to understand the trends, discrepancies. Data has also been checked for any duplicates or null values that might impact the observations and the report. Also, there has been a datatype error in the dataset which is causing the default value of time and date in the fields “Date” and “Time Start” which has been rectified in tableau itself by changing their datatypes to Date and Datetime respectively.

Discrepancy 1: Unidentified Geographic location

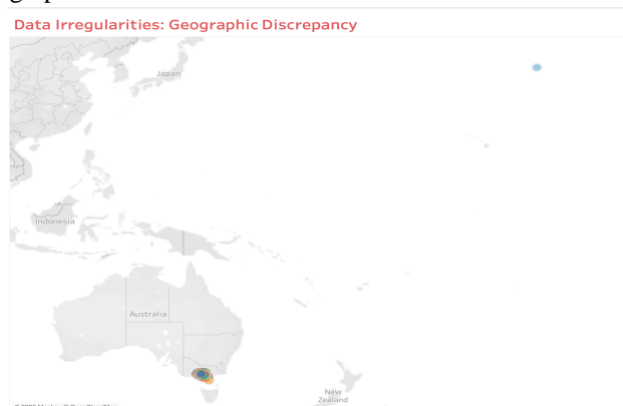


Fig 2: Geographic Locations of Frog Sightings

- The given dataset is for the Port Phillip and Westernport CMA Region which should encompass the area around Victoria but there is one observation that is close to the United States of America.
- Mapping the Unique ID using the Latitude and Longitude gives us the physical location of the data over the map. Hence, we can clearly point the one outlier in the far north of Australia. The unique id of the same is 26630.
- This error can be considered as an outlier and can be excluded while performing exploratory analysis.

Discrepancy 2: Data outside the given timeline

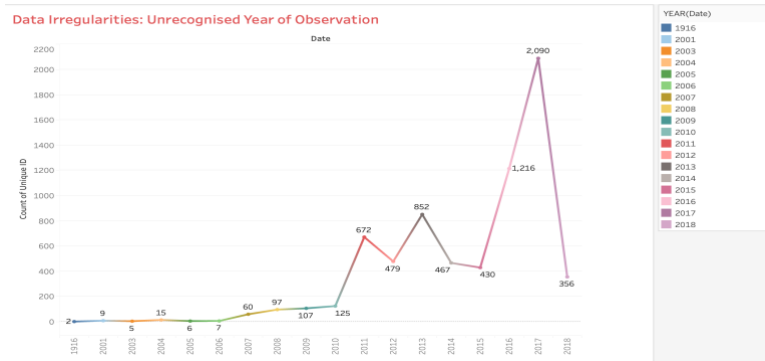


Fig 3: Count of Unique IDs over the years

- The given study is from year 2000-2018.
- Plotting the number of observations made over the years we can clearly see two records for the year 1916 which is not in our timeline of the study making it an outlier. The unique ids 27438 and 27439 are both outliers and can be excluded from the data.

Discrepancy 3: Spelling errors

Pages

Columns

Rows

Family

Scientific name

Genus

Species

Common name

Filters

Scientific name

Marks

Automatic

Color

Size

Text

Detail

Tooltip

CNT(Unique ID)

Sheet 3

Family	Scientific n.	Genus	Species	Common name	
Hylidae	Litoria	Litoria	peronii	Peron's Tree Frog	178
	Litoria		peronella	Peronella's Tree Frog	1
	Litoria	Litoria	raniformis	Growing Grass Frog	75
				Growing grass frog	3

Fig 4: Tabular Data for Spelling Error

- The above figure shows the spelling error in the Common name field.
- This error does impact our analysis as count would be affected as both the spellings will be treated as different values. The unique ids showing these errors are 23837, 23839, 23865 and 33502.
- This can be fixed by making modifications to the data and correcting the value for the name.

After making necessary changes in using Microsoft Excel, we can load the data again to perform our analysis.

Data Visualisation

Q1: 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 behaviour of the researchers or the frogs?

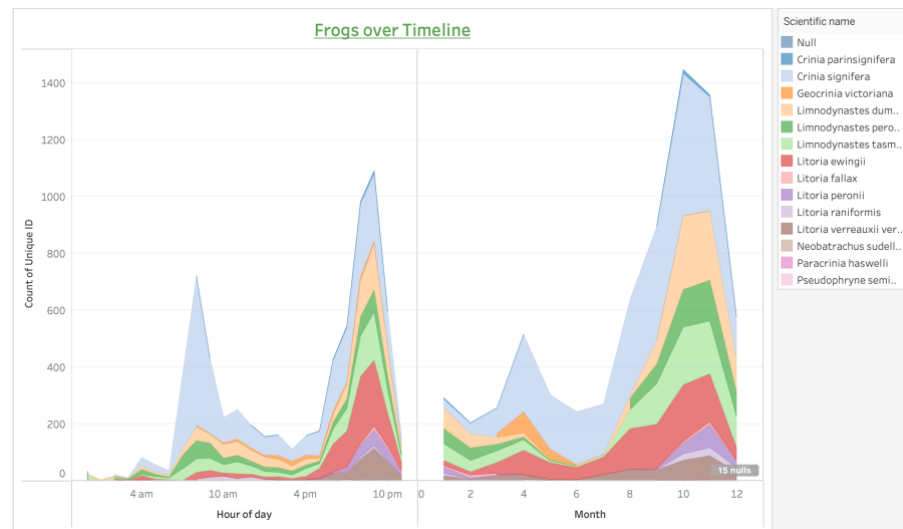


Figure 5: Hourly and Monthly distribution of frogs

- A side-by-side area curve is used to compare the different types of frogs hourly and monthly. Number of Observations i.e., count of unique ids is plotted against the hour of the day and month of the year respectively and colour hue provides us the idea of different types of frogs.
- Line graphs are best to plot timeseries data, but area curves provide more enhanced comparison between different observations.
- The following information can be inferred from the above visualisation:
 - Both monthly and hourly curves show almost similar trend throughout. For hourly observations: It is observed that frog sightings have a peak around at 9 in the morning and 8 at night. Late night observations might suggest that frogs are more active during late hours which might explain their nocturnal behaviour. Also, this can be noted the null value for type of observation which might be due to inactivity of the researchers during these late hours.
 - For monthly observations: There is a dip in the frog sightings around June where Victoria experiences extreme cold weather. Frogs tend to become dormant during the winters which does explains this dip. Also, the peak at around October does suggests that frogs are most likely to active during spring and relatively warmer weather.

Q2: 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?

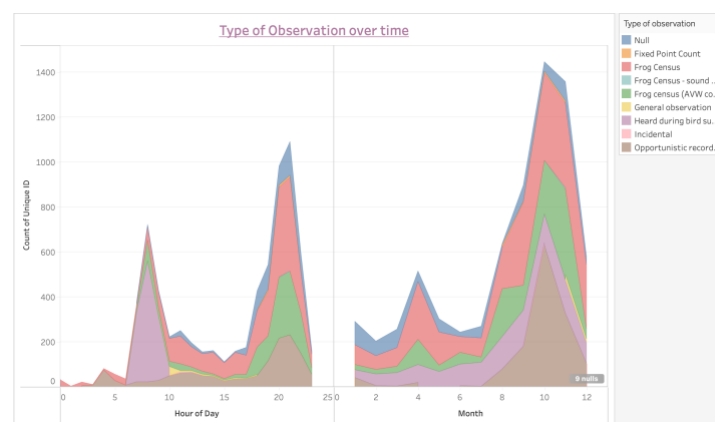


Figure 6: Hourly and Monthly distribution of Type of Observations

-
- Here again, a side-by-side area curve has been used to compare the timeseries data for different values for type of observations. Number of Observations i.e., count of unique ids is plotted against the hour of the day and month of the year respectively and colour hue provides us the idea of different types of observations.
 - The following information can be inferred from the above graphs:
 - ◆ Early morning frog sightings have a peak and the maximum number of observations during this time is recorded by hearing during the bird survey. This explains the trend that we saw in Q1 as frogs are one of the prey for the birds and which might explain their activity during this hour to avoid being hunted. During the night hours again, there is a peak in frog observations which is recorded by Frog Census. This does explain the nocturnal behaviour of the frogs.
 - ◆ For monthly observation: Frog sightings sees a peak during later time in the year and most observations are from opportunistic readings that explains that the frogs are more active during the warmer weather. Sightings due to bird survey is comparatively lesser here as this is the time for bird migration and makes them less susceptible to be hunted.

Assumptions

- The null values are not considered for any fields other than Type of Observation. It is filtered within tableau itself.
- The outliers within the data mentioned before in the discrepancies 1 and 2 are omitted while performing the analysis.
- The spelling errors are rectified by manually editing the values in excel.