Alpha Team

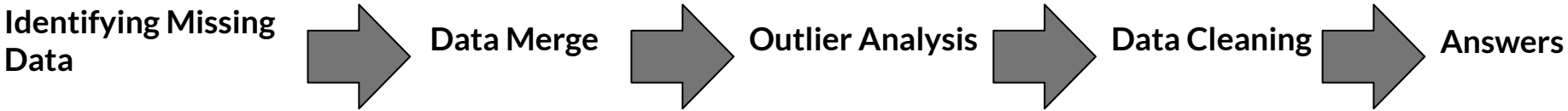
Leeds Data Science Society Data Challenge

Tanmay Jain
Data Enthusiast

Study Background

1. Which trees do the monkeys prefer for rain shelter?
2. Do the monkeys have any fruits they prefer to eat over others?
3. Do the monkeys have any flowers they prefer to eat over others?
4. Are there any trees that the monkeys do not seem to like the fruit of?
5. Are there any trees that the monkeys do not seem to like the flowers of?

Analysis Flow



Data Processing Issues

Missing Data - Errors - Additions

- Monkey count for 300 days missing.
- Rainfall for 40 days missing.
- Change Rainfall & Monkey count Date format .
- Change Monkey count from ‘float’ type to ‘int’ type.
- Tree column has to be converted into Category type for merging
- Introducing Dummy Variables

Solutions

Monkey count 300 days missing

How- Imputing Missing Data/ Deleting Missing Data

Why- To improve accuracy

Before

```
In [3]: monkey_count.isnull().sum()

Out[3]: Date       0
        Tree       0
        count    300
        dtype: int64
```

After

```
monkey_count.isnull().sum()

Date      0
Tree      0
count     0
dtype: int64
```

Rainfall 40 days missing

How- Imputing Missing Data/ Deleting Missing Data

Why- To improve accuracy

Before

```
rain.isnull().sum()
```

```
Date          0
Rainfall     40
dtype: int64
```

After

```
rain.isnull().sum()
```

```
Date        0
Rainfall    0
dtype: int64
```

Change Rainfall & MC date format

How- monkey_count['Date'] = pd.to_datetime(monkey_count['Date'])
rain['Date'] = pd.to_datetime(rain['Date'])

Why- To merge two tables

Before

```
rain.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 2 columns):
Date        731 non-null object
Rainfall    731 non-null float64
dtypes: float64(1), object(1)
memory usage: 11.5+ KB
```

After

```
rain.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 2 columns):
Date        731 non-null datetime64[ns]
Rainfall    731 non-null float64
dtypes: datetime64[ns](1), float64(1)
memory usage: 11.5 KB
```

Change MC ‘float’ to ‘int’ type

How- monkey_count['count'] = monkey_count['count'].astype(int)

Why- MC cannot be in decimals

Before

```
monkey_count.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3500 entries, 0 to 3499
Data columns (total 3 columns):
Date     3500 non-null object
Tree     3500 non-null int64
count    3500 non-null float64
dtypes: float64(1), int64(1), object(1)
memory usage: 82.2+ KB
```

After

```
monkey_count.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3500 entries, 0 to 3499
Data columns (total 3 columns):
Date     3500 non-null datetime64[ns]
Tree     3500 non-null category
count    3500 non-null int64
dtypes: category(1), datetime64[ns](1), int64(1)
memory usage: 58.6 KB
```

Tree column ‘Category Type’

How- monkey_count['Tree'] = monkey_count['Tree'].astype('category')

Why- To merge two tables

Before

```
monkey_count.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3500 entries, 0 to 3499
Data columns (total 3 columns):
Date     3500 non-null object
Tree     3500 non-null int64
count    3500 non-null float64
dtypes: float64(1), int64(1), object(1)
memory usage: 82.2+ KB
```

After

```
monkey_count.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3500 entries, 0 to 3499
Data columns (total 3 columns):
Date     3500 non-null datetime64[ns]
Tree     3500 non-null category
count    3500 non-null int64
dtypes: category(1), datetime64[ns](1), int64(1)
memory usage: 58.6 KB
```

Merged Data

```
final_data.head()
```

	Date	Tree	count	Rainfall	Tree species	Flowering	Fruiting	Average leaf diameter	Average leaf coverage
0	2016-01-01	1	4	0.824823	Red bucket tree	January-February	March	2	84
1	2016-01-02	1	4	0.619122	Red bucket tree	January-February	March	2	84
2	2016-01-03	1	2	2.055951	Red bucket tree	January-February	March	2	84
3	2016-01-04	1	4	1.156919	Red bucket tree	January-February	March	2	84
4	2016-01-05	1	5	0.460096	Red bucket tree	January-February	March	2	84

```
final_data.shape
```

```
(3500, 9)
```

Introducing ‘Dummy Variables’

How- **Using ‘lambda’ function then using ‘Dummy Variables’**

Why- To convert categorical variables into quantitative variables

Dummy Column

Month_Name

	Date	Tree	count	Rainfall	Tree species	Flowering	Fruiting	Average leaf diameter	Average leaf coverage	Month_Name
0	2016-01-01	1	4	0.824823	Red bucket tree	January-February	March	2	84	January
1	2016-01-02	1	4	0.619122	Red bucket tree	January-February	March	2	84	January
2	2016-01-03	1	2	2.055951	Red bucket tree	January-February	March	2	84	January
3	2016-01-04	1	4	1.156919	Red bucket tree	January-February	March	2	84	January
4	2016-01-05	1	5	0.460096	Red bucket tree	January-February	March	2	84	January

Dummy columns

	Date	Tree	count	Rainfall	Tree species	Flowering	Fruiting	Average leaf diameter	Average leaf coverage	Month_Name	Dummy_Flowering	Dummy_Fruiting
0	2016-01-01	1	4	0.824823	Red bucket tree	[January, February]	[March]	2	84	January	1	0
1	2016-01-02	1	4	0.619122	Red bucket tree	[January, February]	[March]	2	84	January	1	0
2	2016-01-03	1	2	2.055951	Red bucket tree	[January, February]	[March]	2	84	January	1	0
3	2016-01-04	1	4	1.156919	Red bucket tree	[January, February]	[March]	2	84	January	1	0
4	2016-01-05	1	5	0.460096	Red bucket tree	[January, February]	[March]	2	84	January	1	0

Final head of data to be used for analysis

	Date	Tree	count	Rainfall	Tree species	Average leaf diameter	Average leaf coverage	Dummy_Flowering	Dummy_Fruiting
0	2016-01-01	1	4	0.824823	Red bucket tree	2	84	1	0
1	2016-01-02	1	4	0.619122	Red bucket tree	2	84	1	0
2	2016-01-03	1	2	2.055951	Red bucket tree	2	84	1	0
3	2016-01-04	1	4	1.156919	Red bucket tree	2	84	1	0
4	2016-01-05	1	5	0.460096	Red bucket tree	2	84	1	0

Statistical Analysis Methods

- Mean
- Standard Deviation
- Regression

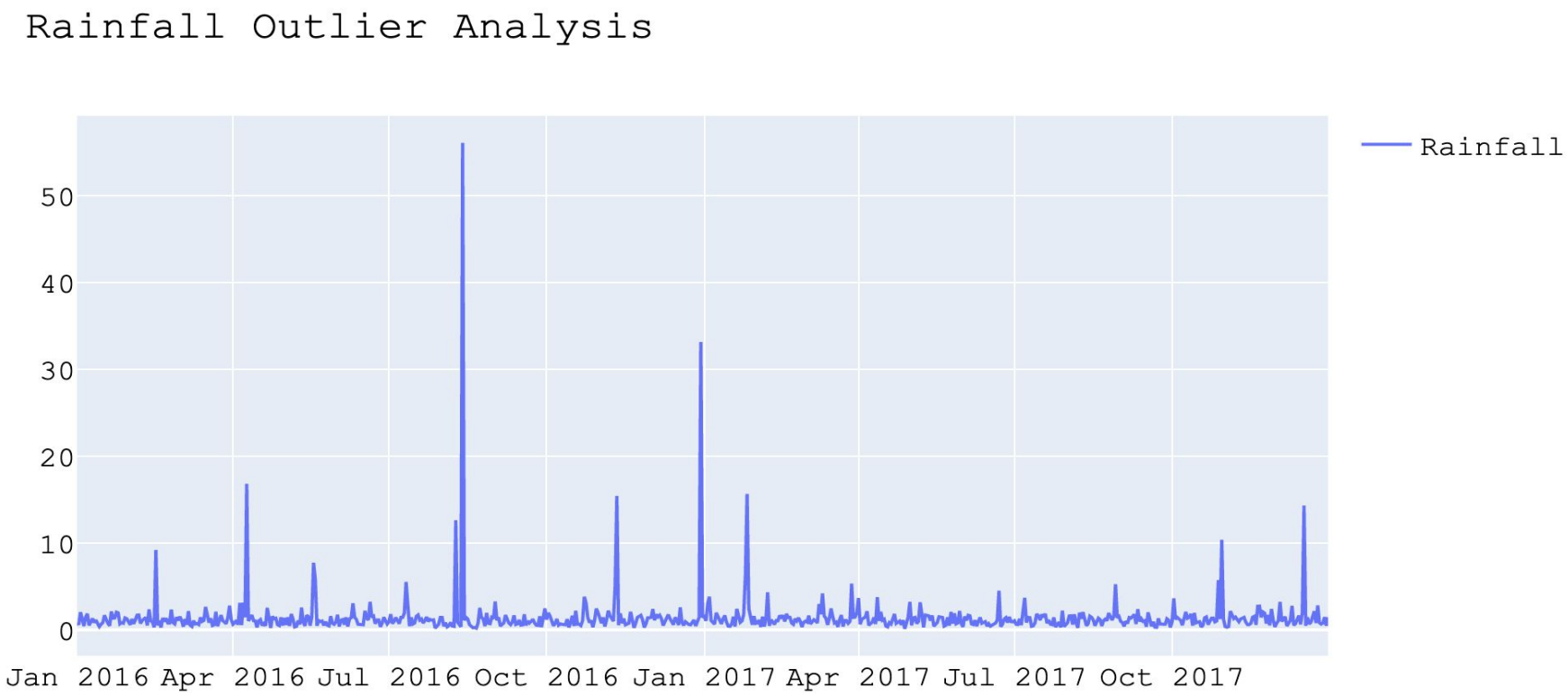
Outlier Analysis

- Monkey count outlier analysis
- Rainfall outlier analysis
- Tree species outlier analysis

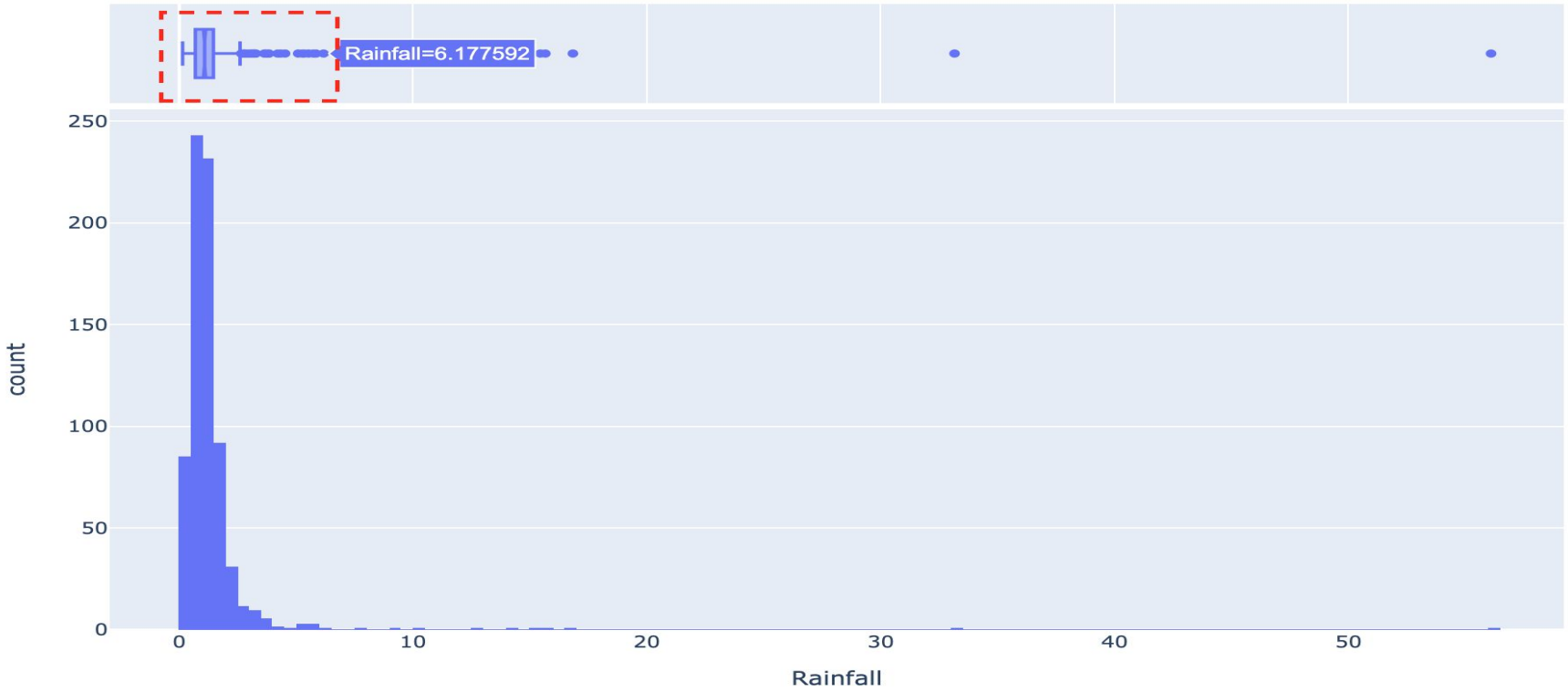
Data Cleaning

- Restricting rainfall to 6.1 cm
- Restricting monkey count to 20

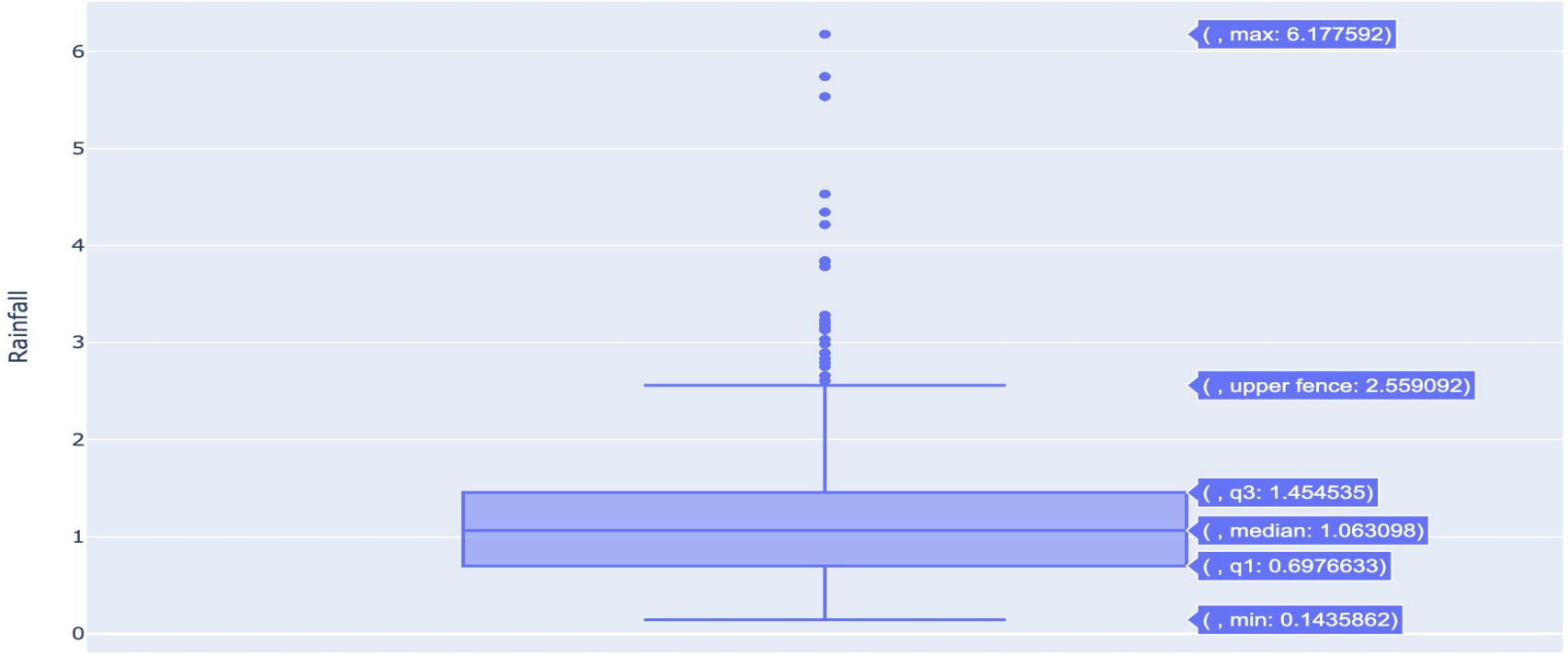
Rainfall outlier analysis



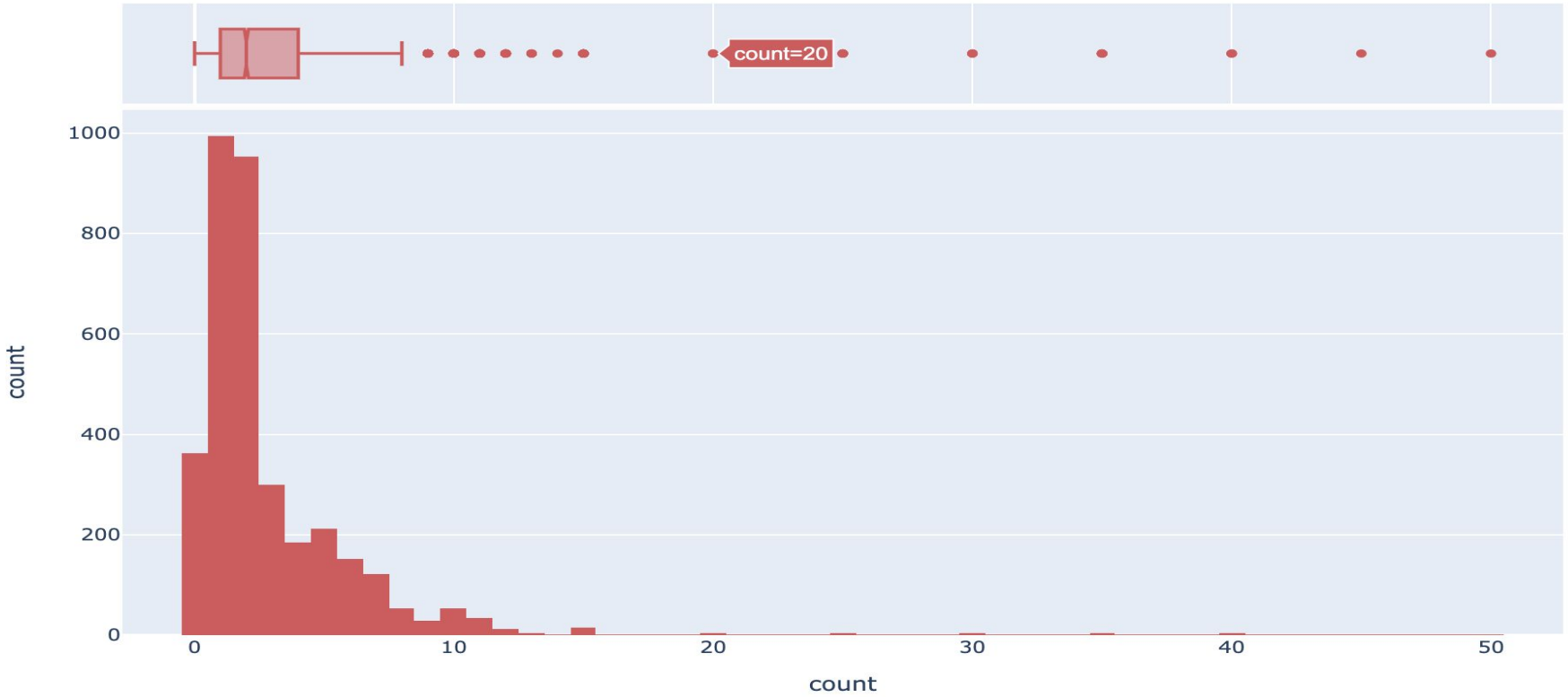
Rainfall outlier analysis box & histogram



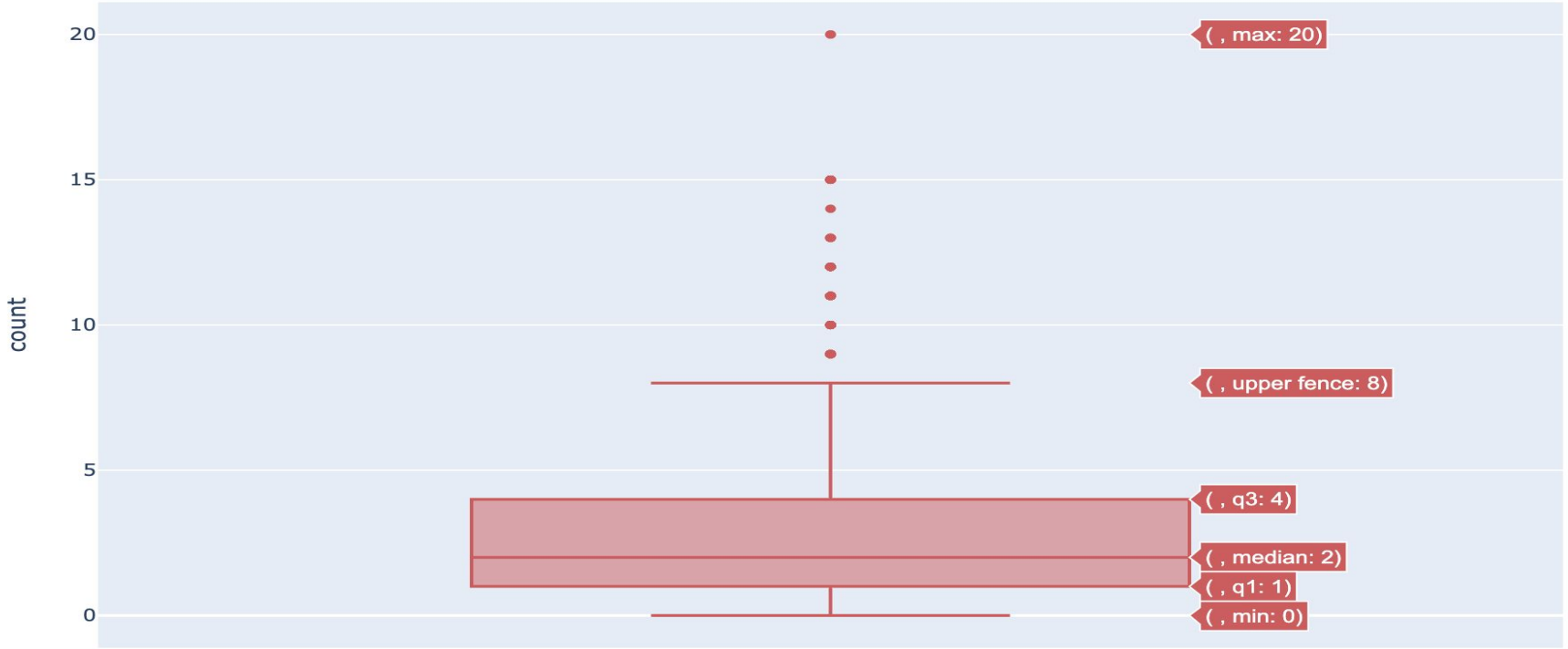
Final rainfall data for analysis



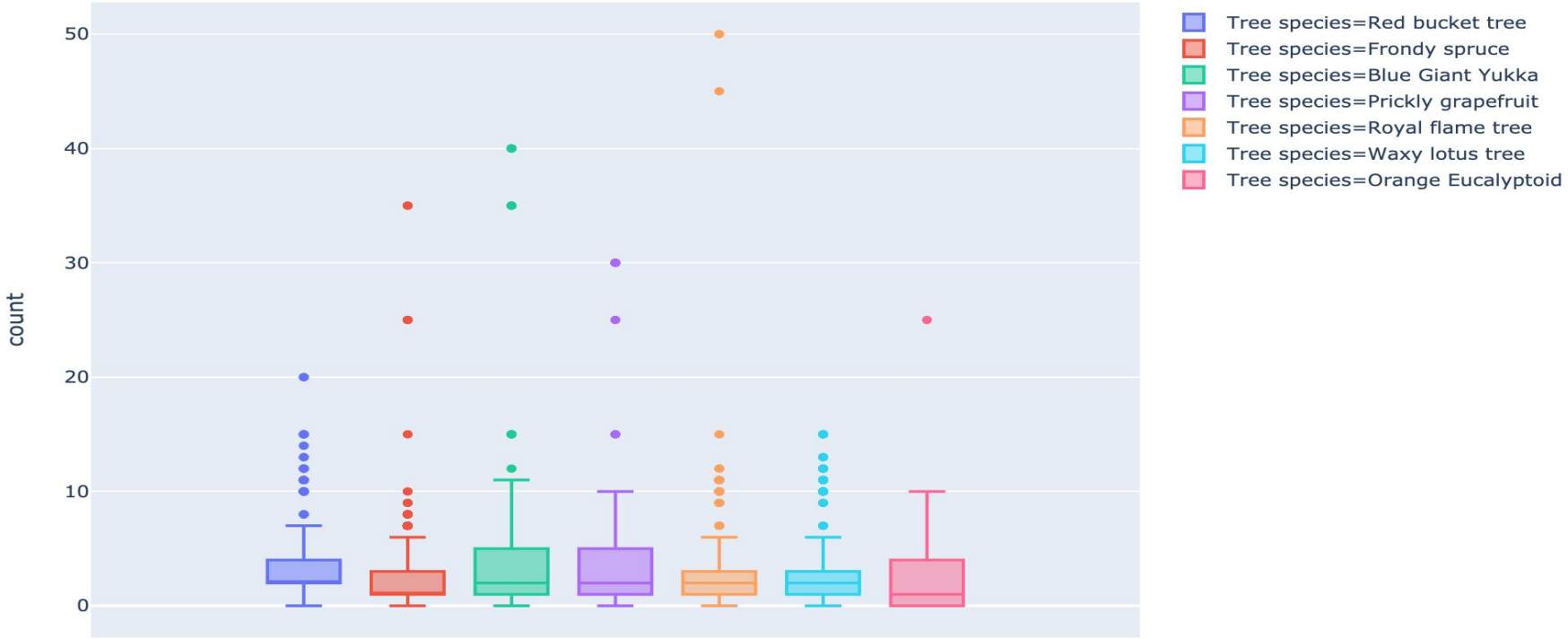
Monkey count outlier analysis



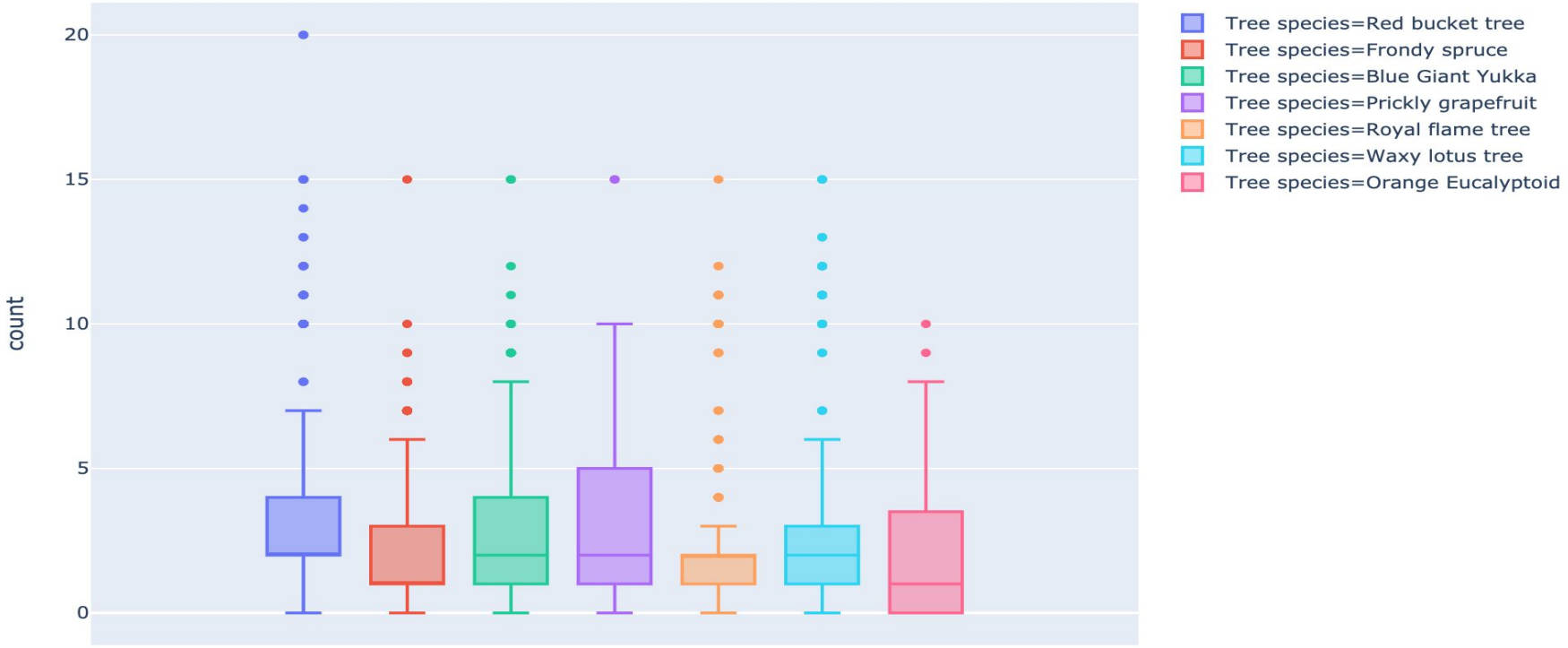
Final monkey count data for analysis



Monkey count on each tree



Monkey count restricted to 20 for analysis



Answers
&
Conclusion

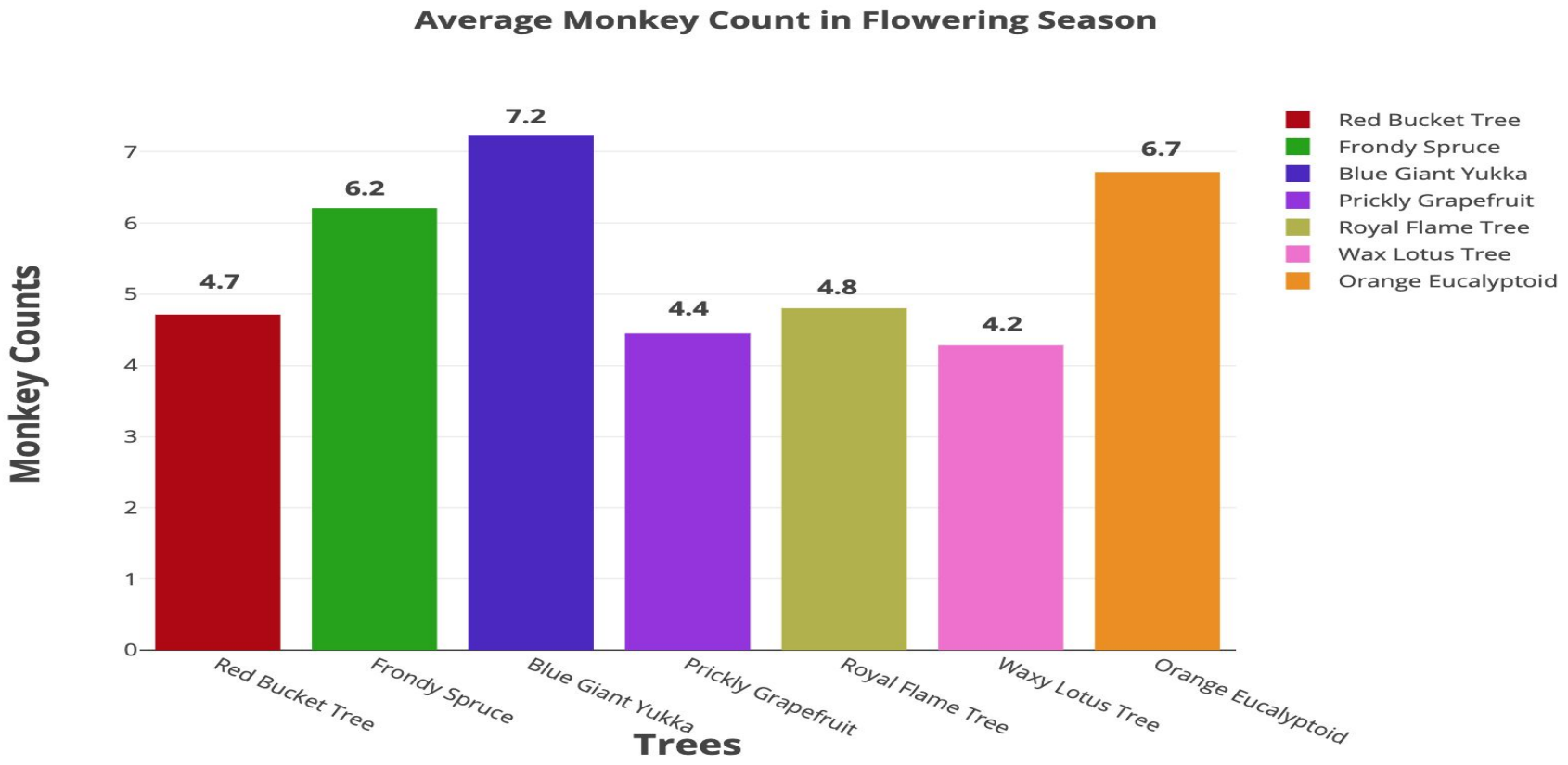
Do the monkeys have any flowers they prefer to eat over others?

1. Blue Giant Yukka- 7

2. Orange Eucalyptoid- 7

Are there any tree that monkeys do not seem to like the flowers of?

1. Waxy Lotus Tree - 4


Average Monkey Count in Flowering Season
Monkey Counts
Trees
0
1
2
3
4
5
6
7
4.7
6.2
7.2
4.4
4.8
4.2
6.7
Red Bucket Tree
Frondy Spruce
Blue Giant Yukka
Prickly Grapefruit
Royal Flame Tree
Waxy Lotus Tree
Orange Eucalyptoid
Red Bucket Tree
Frondy Spruce
Blue Giant Yukka
Prickly Grapefruit
Royal Flame Tree
Wax Lotus Tree
Orange Eucalyptoid

Do the monkeys have any fruits they prefer to eat over others?

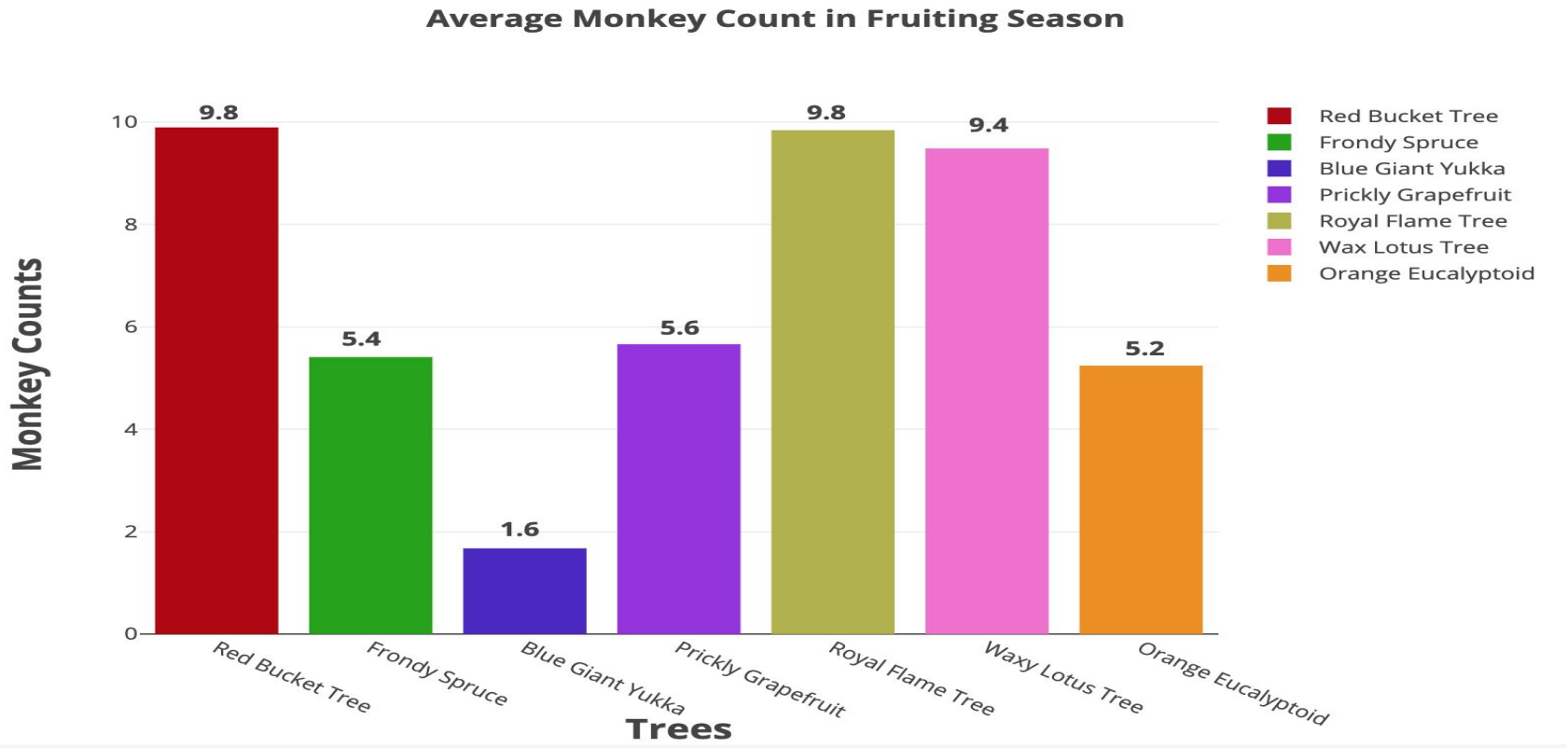
1. Red Bucket Tree & Royal Flame Tree - 10

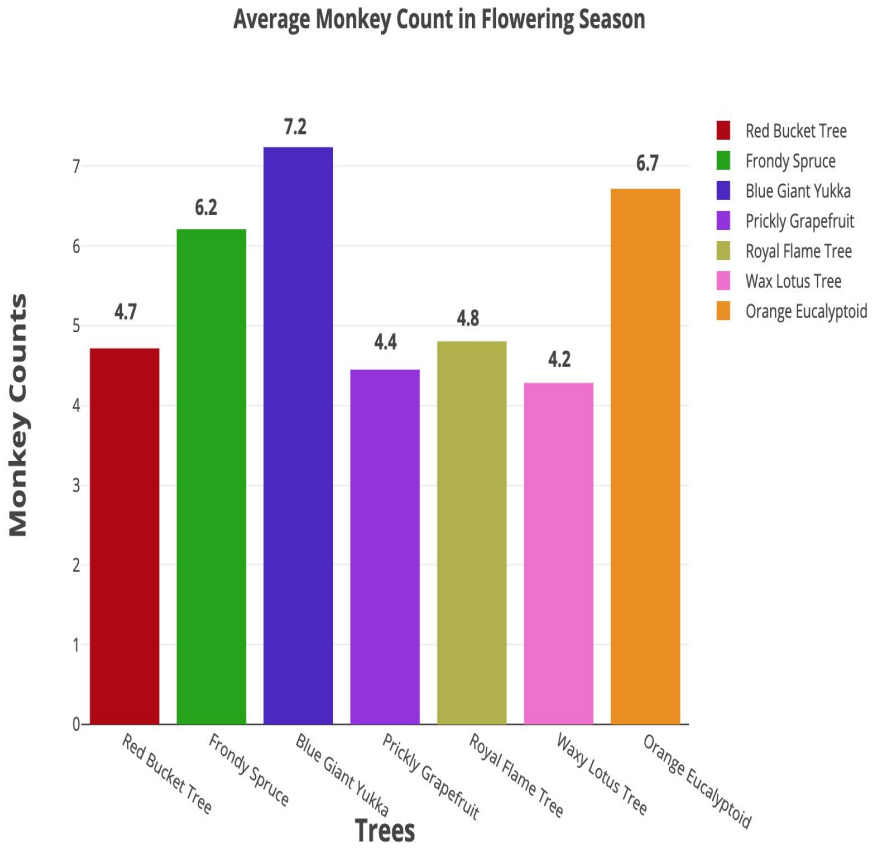
2. Waxy Lotus Tree - 9

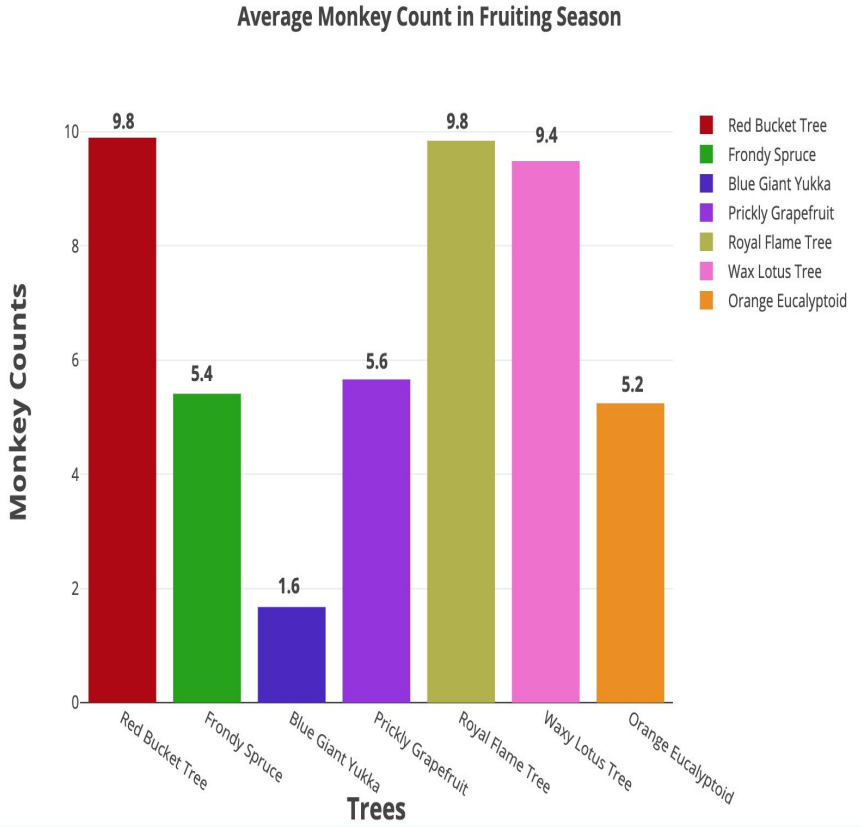
Are there any tree that monkeys do not seem to like the fruit of?

1. Blue Giant Yukka - 2

-


Average Monkey Count in Fruiting Season
Monkey Counts
10
8
6
4
2
0
9.8
5.4
1.6
5.6
9.8
9.4
5.2
Red Bucket Tree
Frondy Spruce
Blue Giant Yukka
Prickly Grapefruit
Royal Flame Tree
Waxy Lotus Tree
Orange Eucalyptoid
Trees



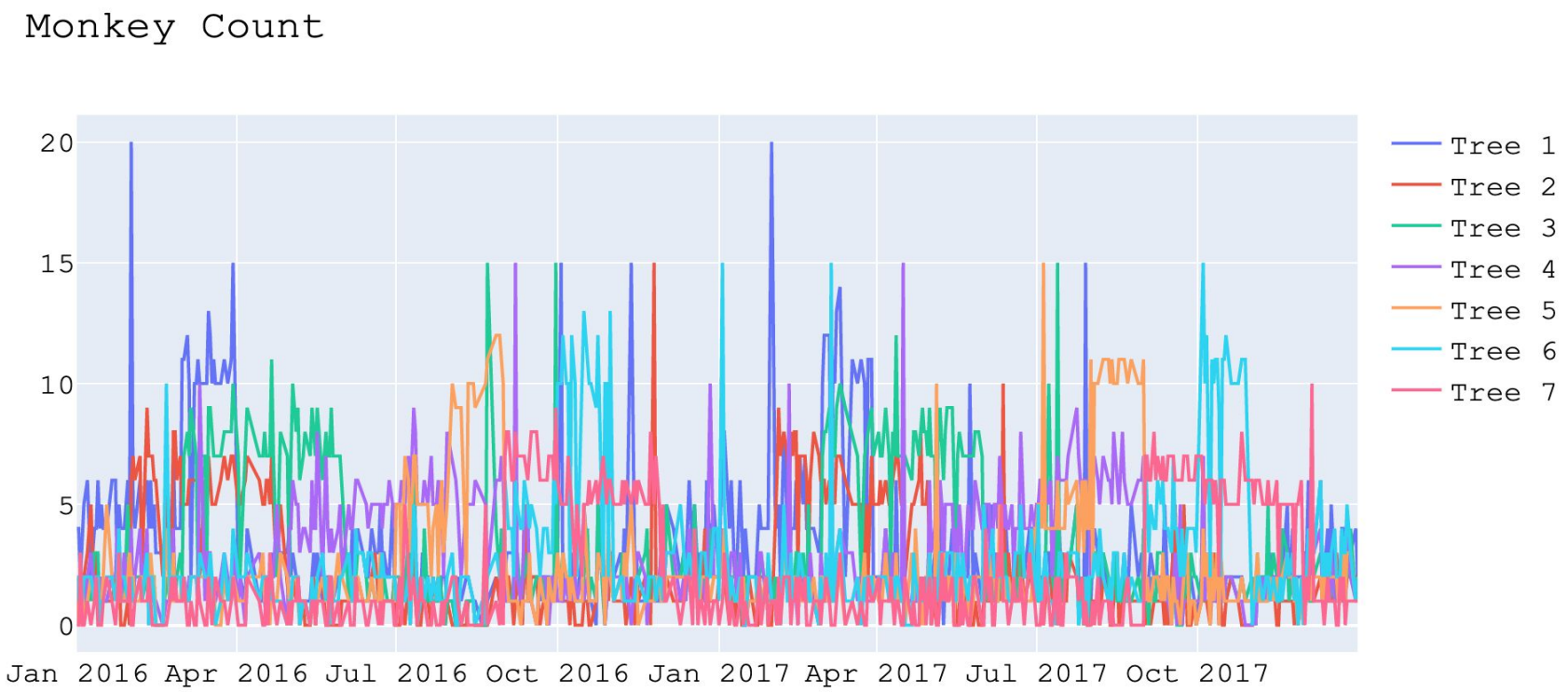


Flowering & Fruiting

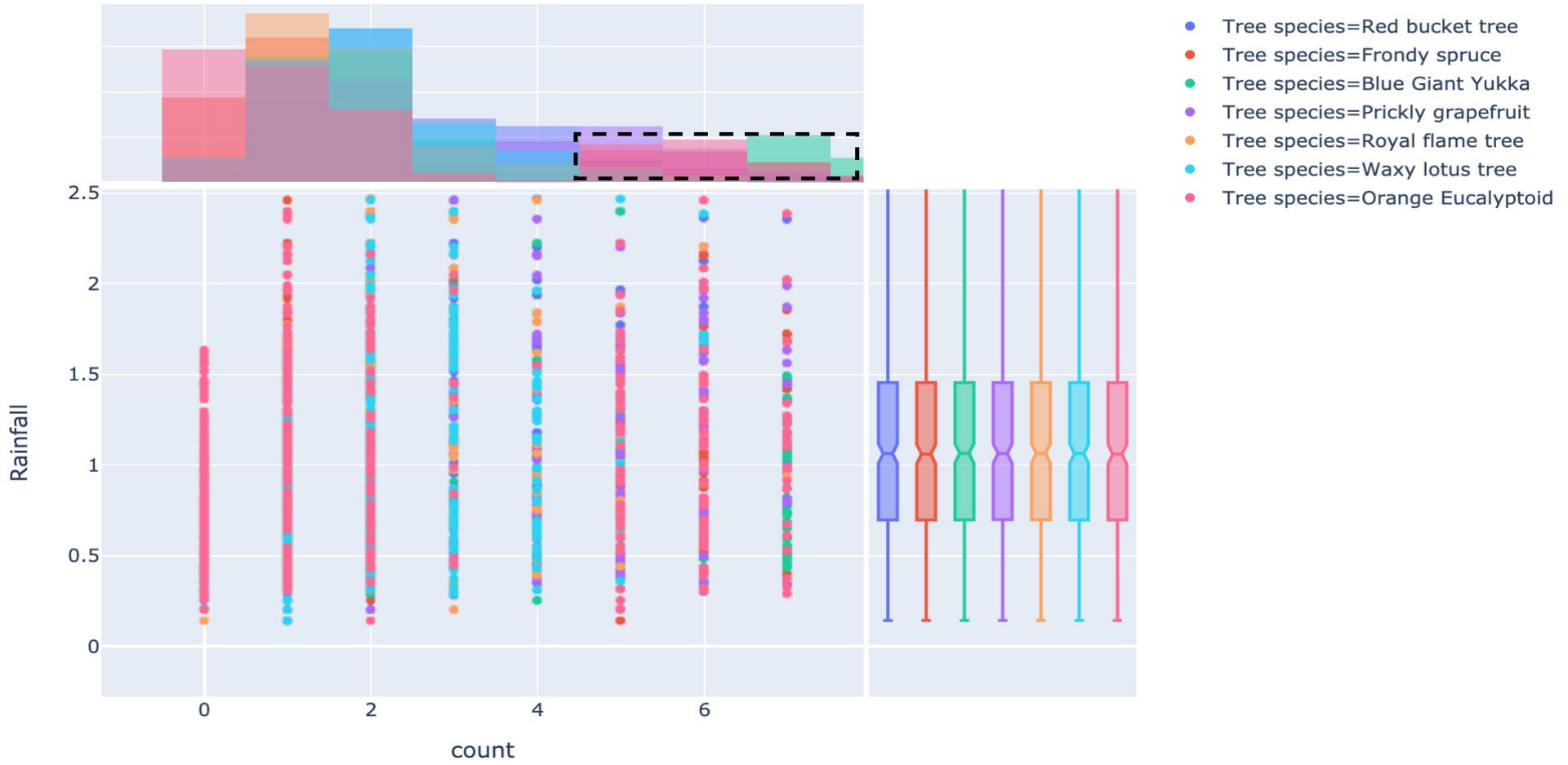
Which trees do the monkeys prefer for rain shelters?

1. Blue Giant Yukka

2. Orange Eucalyptoid

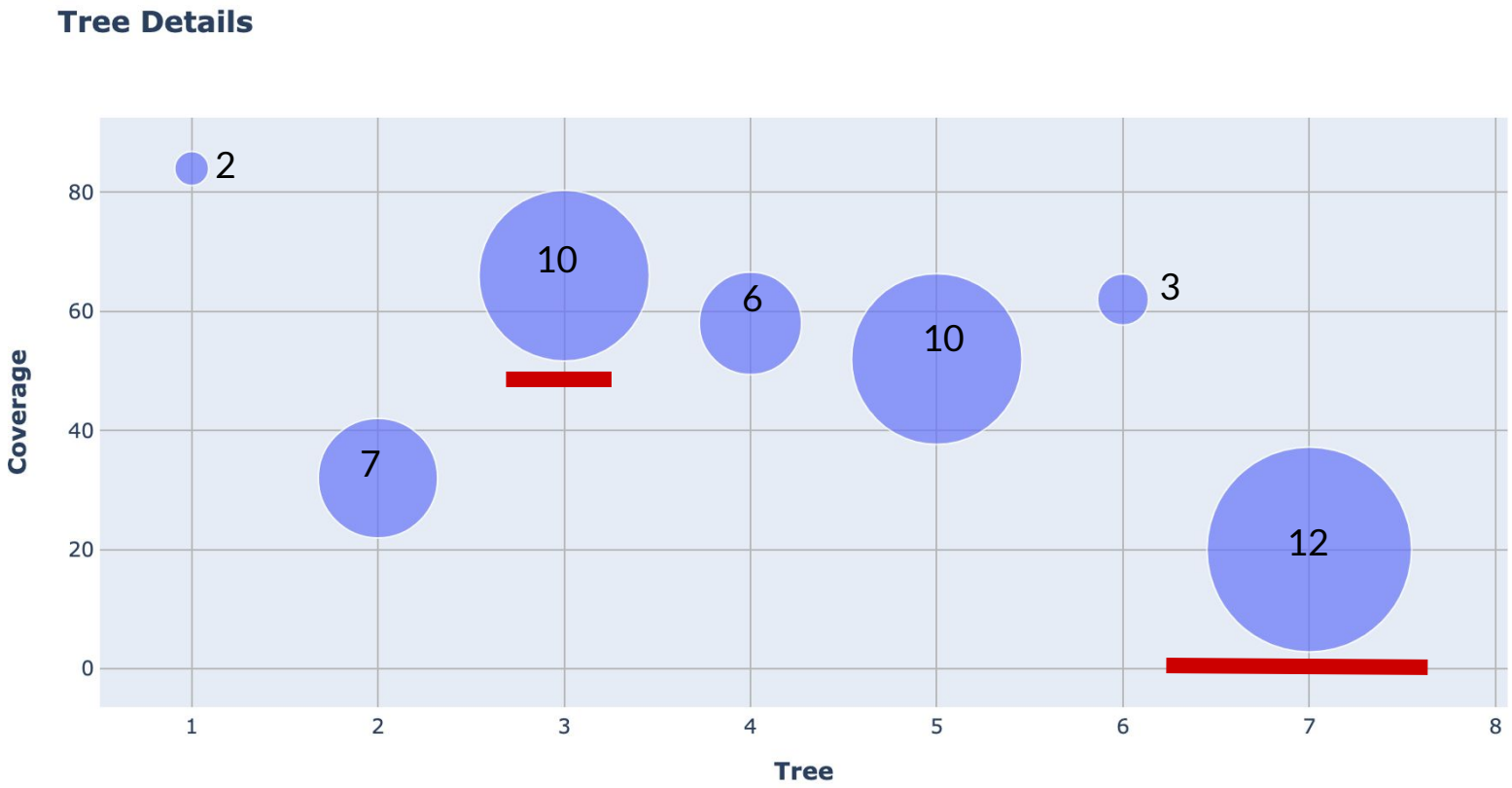


Monkey count on each tree



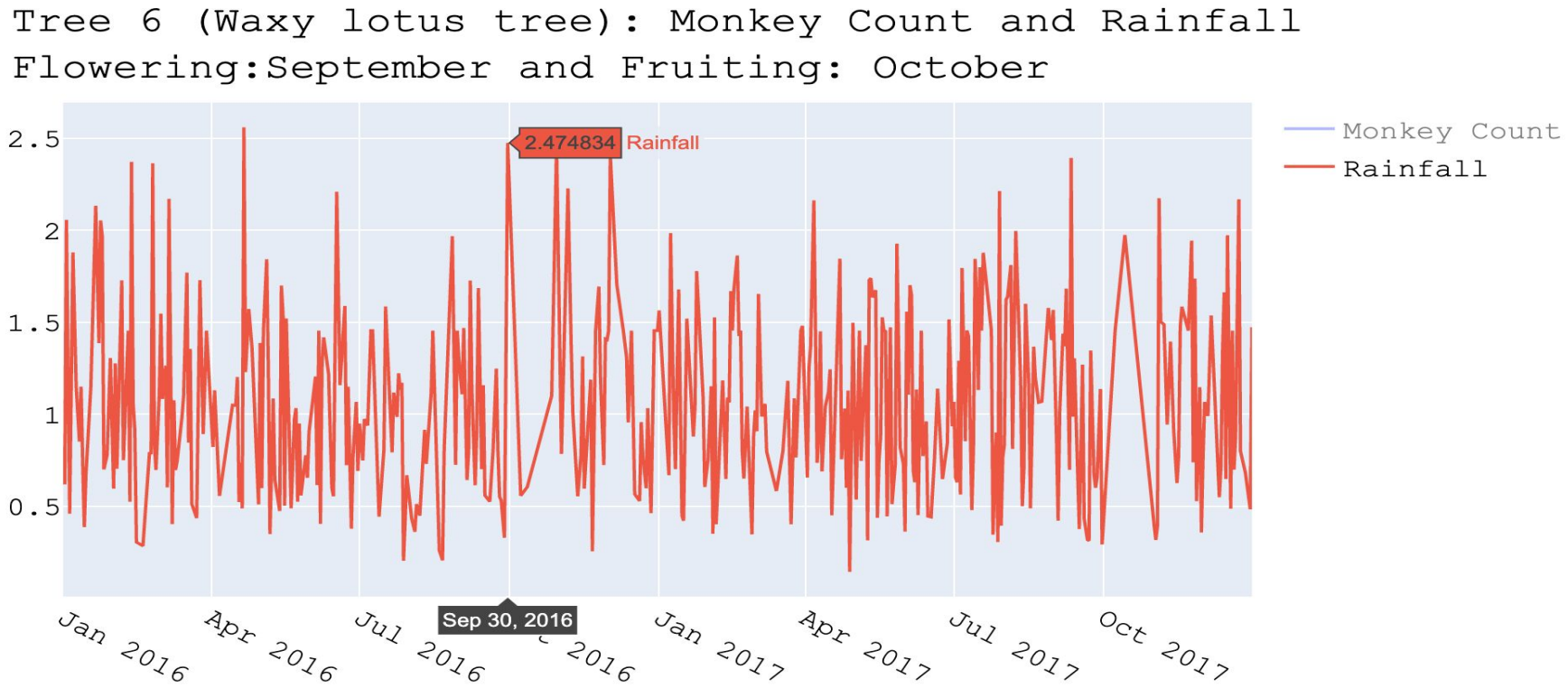
Monkey count during rainfall

Diameter & coverage of tree

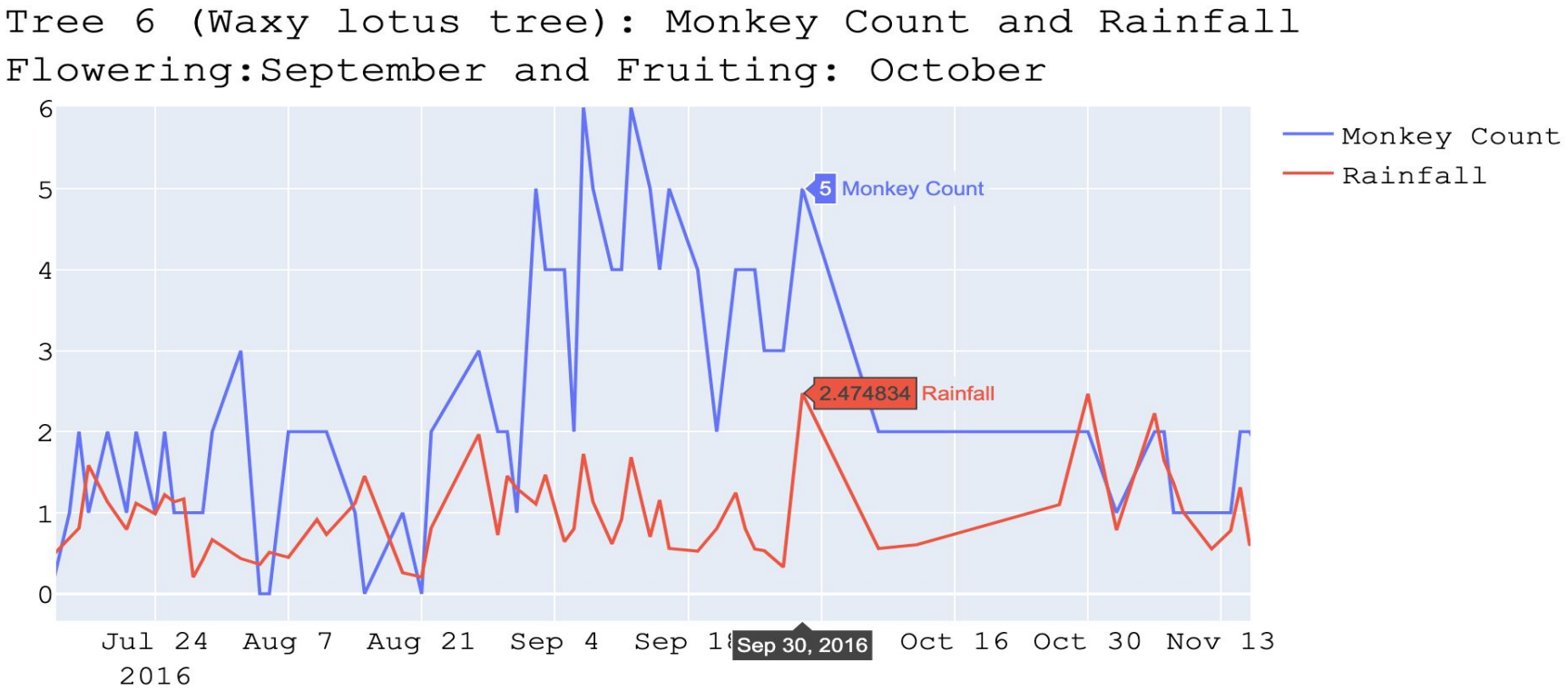


Limitations

Rainfall


Tree 6 (Waxy lotus tree): Monkey Count and Rainfall
Flowering:September and Fruiting: October
2.5
2
1.5
1
0.5
2.474834 Rainfall
Sep 30, 2016
Jan 2016
Apr 2016
Jul 2016
Jan 2017
Apr 2017
Jul 2017
Oct 2017
Monkey Count
Rainfall

Reason for higher monkey count unknown- Eating Fruits- Shelter- Eating leaves. Don't like flowers



Thank You

Questions?