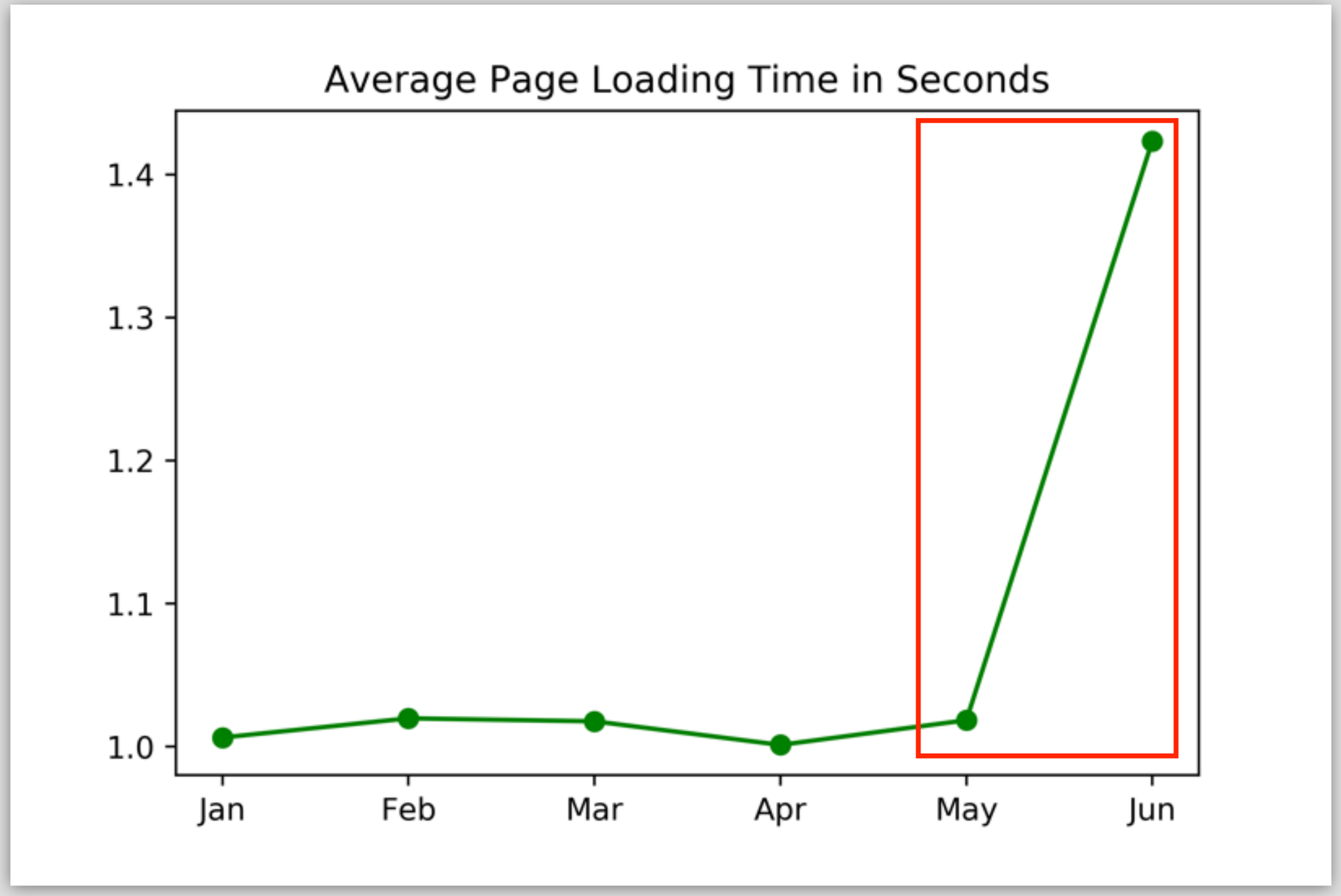
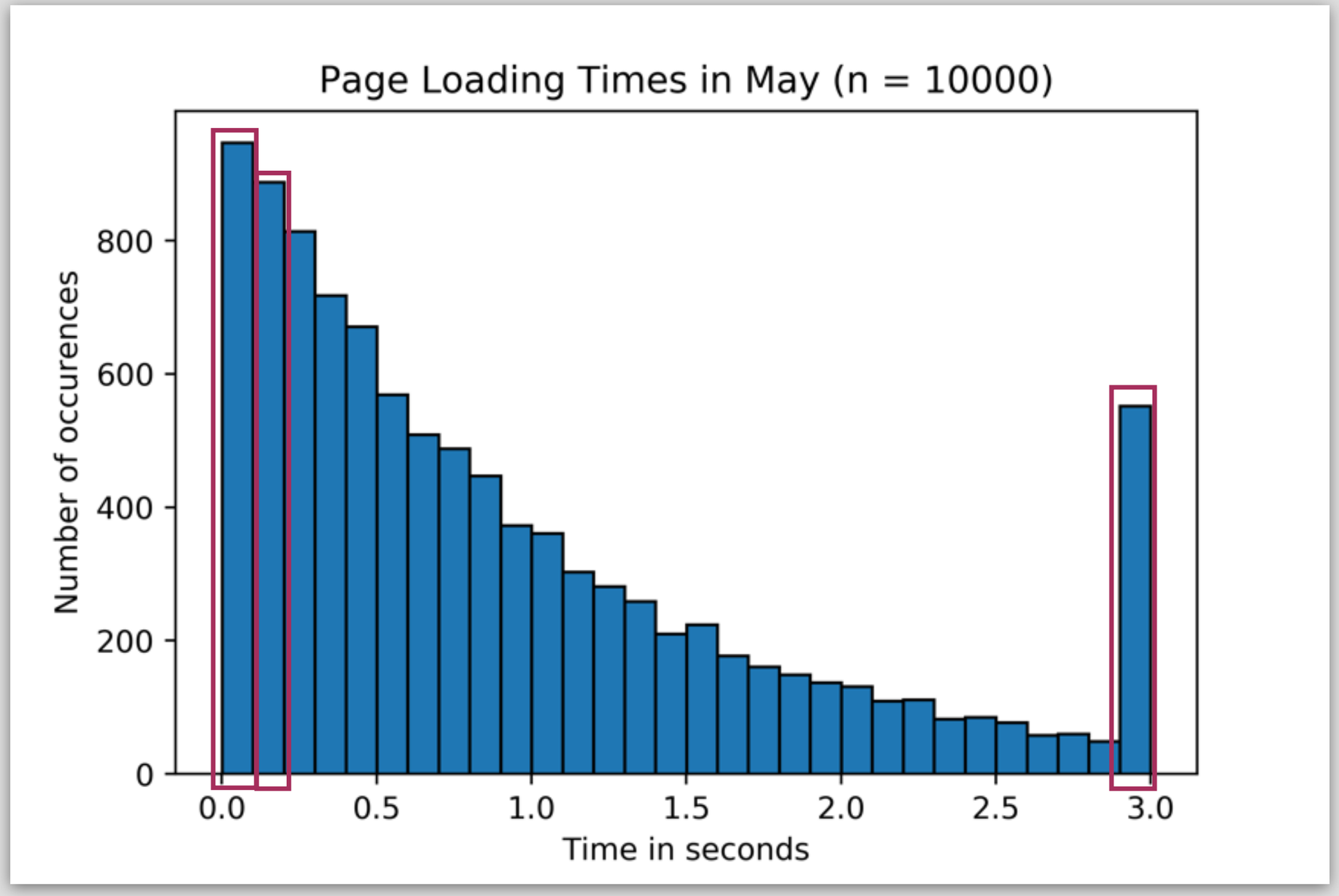
# Finding Distribution of Data with Histograms

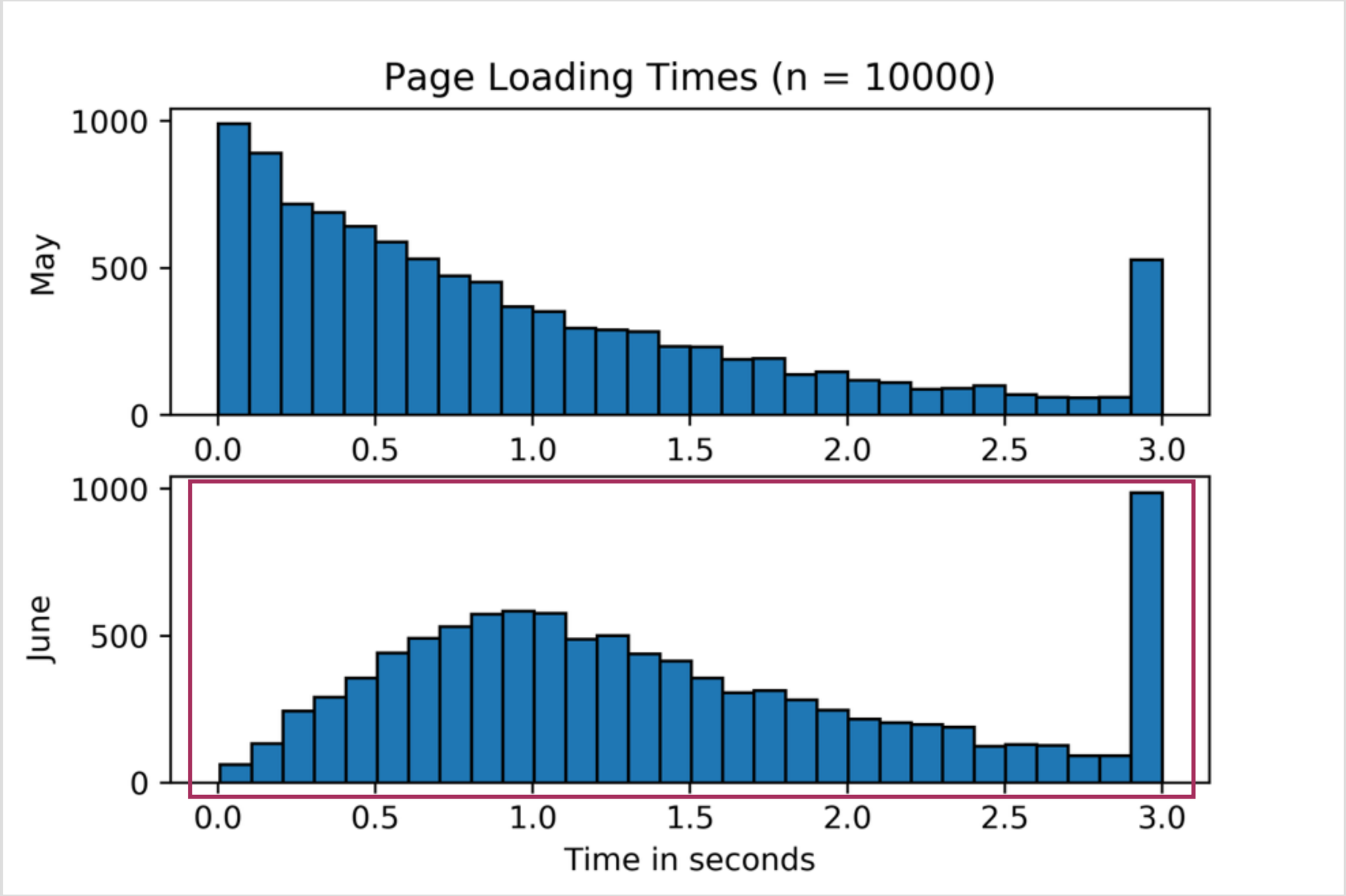
**YK Sugishita**
SOFTWARE DEVELOPER / DATA SCIENTIST
www.csdojo.io

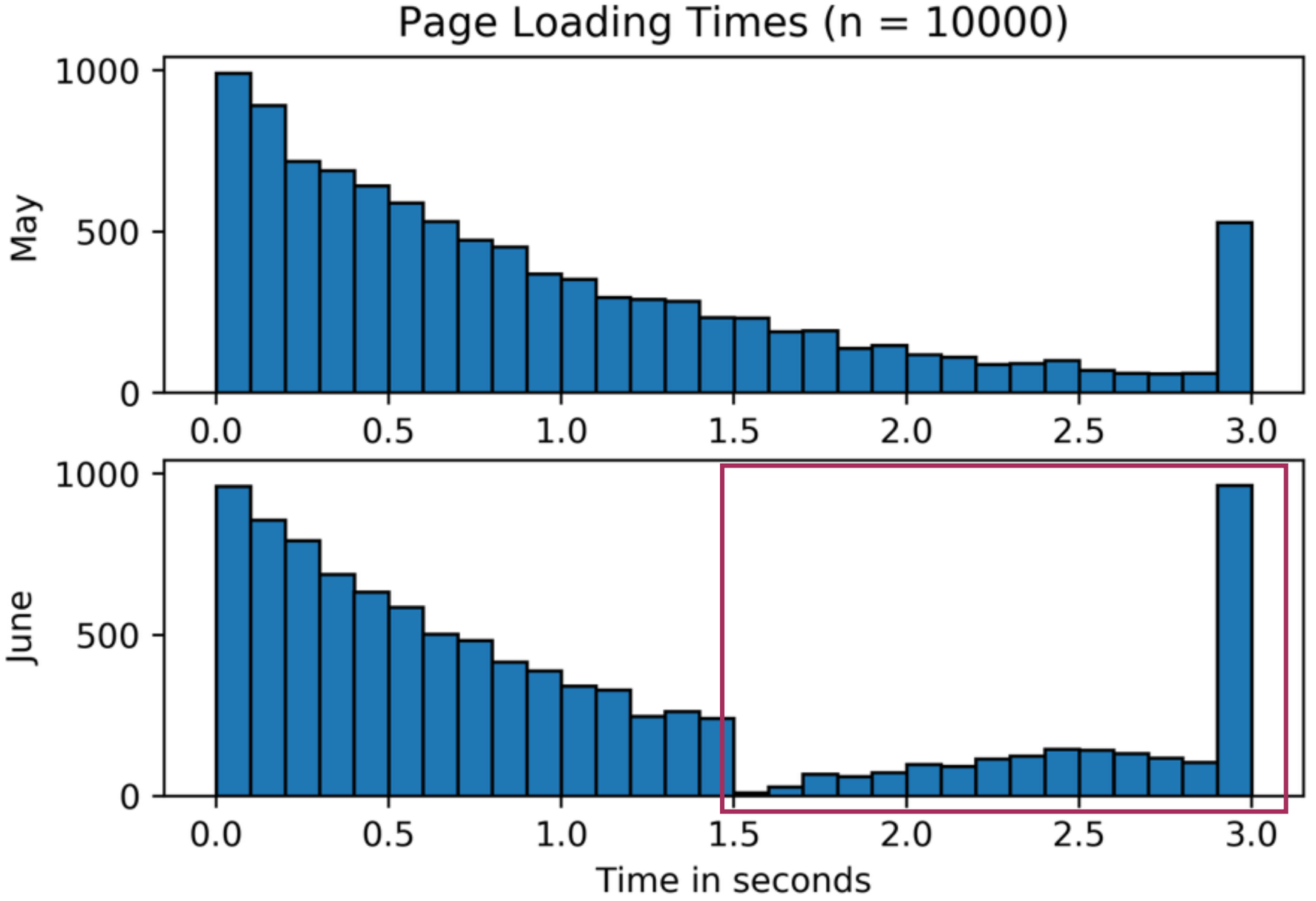

A3
B3
A4
B4


Average Page Loading Time in Seconds
1.4
1.3
1.2
1.1
1.0
Jan
Feb
Mar
Apr
May
Jun




**Note: Any value larger than 3 is bundled into the last bin**


Page Loading Times (n = 10000)
May
June
1000
500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time in seconds


Page Loading Times (n = 10000)
May
June
1000
500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time in seconds

# Histograms

Helps you understand the distribution of a numeric value in a way that you cannot with mean or median alone

# Creating Histograms with Matplotlib

# Problem: Comparing Asia and Europe's GDP Per Capita in 2007

# countries.csv

| country | continent | year | lifeExpectancy | population | gdpPerCapita |
|---|---|---|---|---|---|
| Afghanistan | Asia | 1952 | 28.801 | 8425333 | 779.4453145 |
| Afghanistan | Asia | 1957 | 30.332 | 9240934 | 820.8530296 |
| Afghanistan | Asia | 1962 | 31.997 | 10267083 | 853.10071 |
| | | | ⋮ | | |
| Zimbabwe | Africa | 1997 | 46.809 | 11404948 | 792.4499603 |
| Zimbabwe | Africa | 2002 | 39.989 | 11926563 | 672.0386227 |
| Zimbabwe | Africa | 2007 | 43.487 | 12311143 | 469.7092981 |

# Demo

**Creating Histograms with Matplotlib**

- Create histograms
- Use Matplotlib's subplot() function

# Practice Problem 1 - Histograms

# Problem: Compare Europe and Americas' Life Expectancy in 1997

# Compare Europe and Americas' Life Expectancy

Use Matplotlib's hist() function

Use subplot() to compare histograms

Pause video

# Demo

## Example solution