

A3 Recap, Plotting, A4

Learning goals:

- Go over commonly asked questions for A3.
- Understand how common Python plotting libraries relate with each other.
- Walk through questions on A4

COGS 108 Winter 2020
Amir Farhan
Discussion 6

bit.ly/sam-wi20
amatkama@ucsd.edu
OH: TBA

Plotting

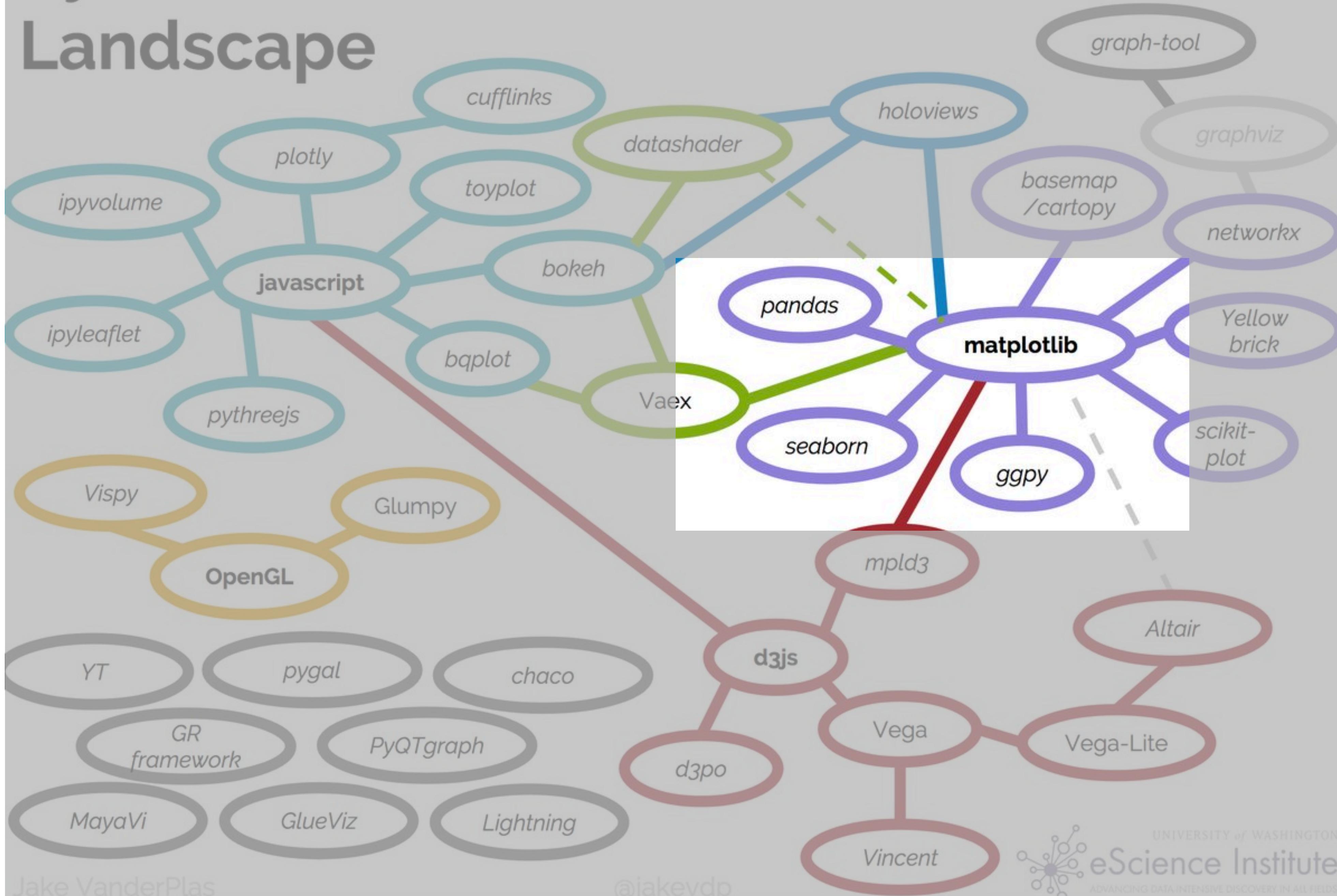
Why are there so many ways to make the same plot?

- All of these do the same thing:

```
plt.hist(df['income10'], 25)  
df['income10'].hist(bins=25)  
df.hist('income10', bins=25)
```

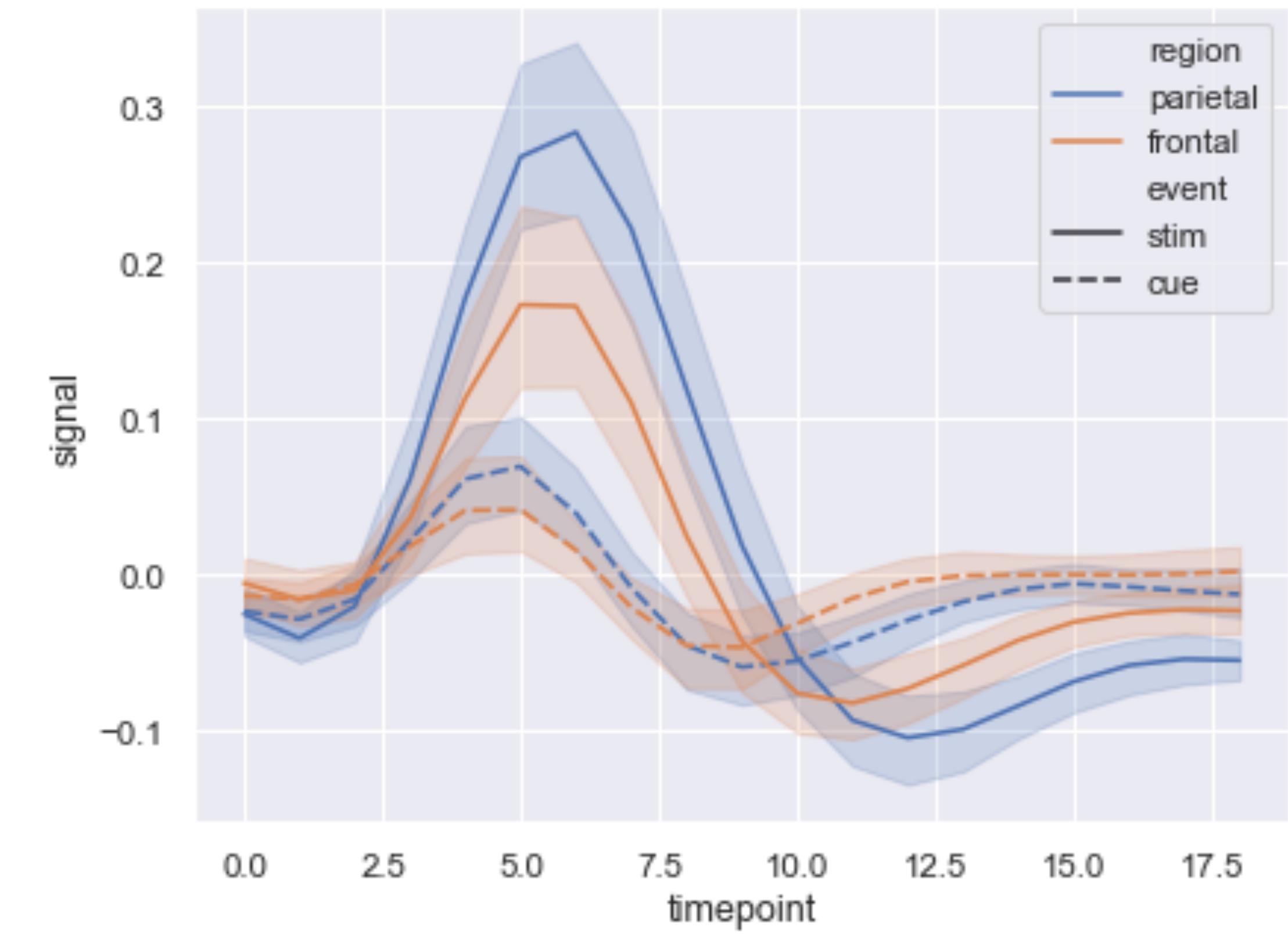
- In Python, most image-based plots created using Matplotlib.
 - plt.hist plt.bar plt.plot etc.
- Pandas gives shortcuts for matplotlib plots. Lines 2 and 3 are shortcuts for line 1.

Python's Visualization Landscape



Seaborn

- **My personal favorite is the seaborn library.**
- **Makes common statistical charts easy to create, like bar plots with confidence intervals.**
- **Again, seaborn is really just a bunch of shortcuts for matplotlib.**



For more details

- **Making good plots is a key skill! This just scratches surface.**
- **You can get many great jobs just by being able to make informative data visualizations.**
- **For more, see Ch 6 of [textbook.ds100.org](#).**

A4 Walkthroughs

Preview of next week

An easy way to set up a personal website using Jupyter notebooks and GitHub.

A5 question walkthroughs

The Waiting Time Paradox, or, Why Is My Bus Always Late?

Thu 13 September 2018



Image Source: [Wikipedia](#) License CC-BY-SA 3.0

If you, like me, frequently commute via public transit, you may be familiar with the following situation:

You arrive at the bus stop, ready to catch your bus: a line that advertises arrivals every 10 minutes. You glance at your watch and note the time... and when the bus finally comes 11 minutes later, you wonder why you always seem to be so unlucky.

Naïvely, you might expect that if buses are coming every 10 minutes and you arrive at a random time, your average wait would be something like 5 minutes. In reality, though, buses do not arrive exactly on schedule, and so you might wait longer. It turns out that under some reasonable assumptions, you can reach a startling conclusion:

When waiting for a bus that comes on average every 10 minutes, your average waiting time will be 10 minutes.

This is what is sometimes known as the *waiting time paradox*.

I've encountered this idea before, and always wondered whether it is actually true... how well do those "reasonable assumptions" match reality? This post will explore the waiting time paradox from the standpoint of both simulation and probabilistic arguments, and then take a look at some real bus arrival time data from the city of Seattle to (hopefully) settle the paradox once and for all.

The Inspection Paradox

If buses arrive exactly every ten minutes, it's true that your average wait time will be half that interval: 5 minutes. Qualitatively speaking, it's easy to convince yourself that adding some variation to those arrivals will make the average wait time somewhat longer, as we'll see here.

The waiting time paradox turns out to be a particular instance of a more general phenomenon, the *inspection paradox*, which is discussed at length in this enlightening post by Allen Downey: [The Inspection Paradox Is Everywhere](#).

Briefly, the inspection paradox arises whenever the probability of observing a quantity is related to

A4 quick tips

- **1d: Your DF cells should have `\\n` at the end (same for 1e)**
- **2b: Don't manually make a new Series – slice a column out of a DF**
- **2e: Use a slice with multiple boolean expressions**
- **2f: Your for loop should loop through the index of a DF**
- **3b: Don't do anything to the zip column**
- **3f: Loop through the DF's index again. Your 3-digit zip codes should be stored as strings, not ints.**
- **Attendance: bit.ly/at-wi20**