Data Viz with Seaborn

Learning goals:

- Understand the basics of tidy data.
- Learn how seaborn enables fast exploration of data.
- Make progress on A4.

COGS 108 Fall 2019

Sam Lau

Discussion 8

bit.ly/sam-108-fa19

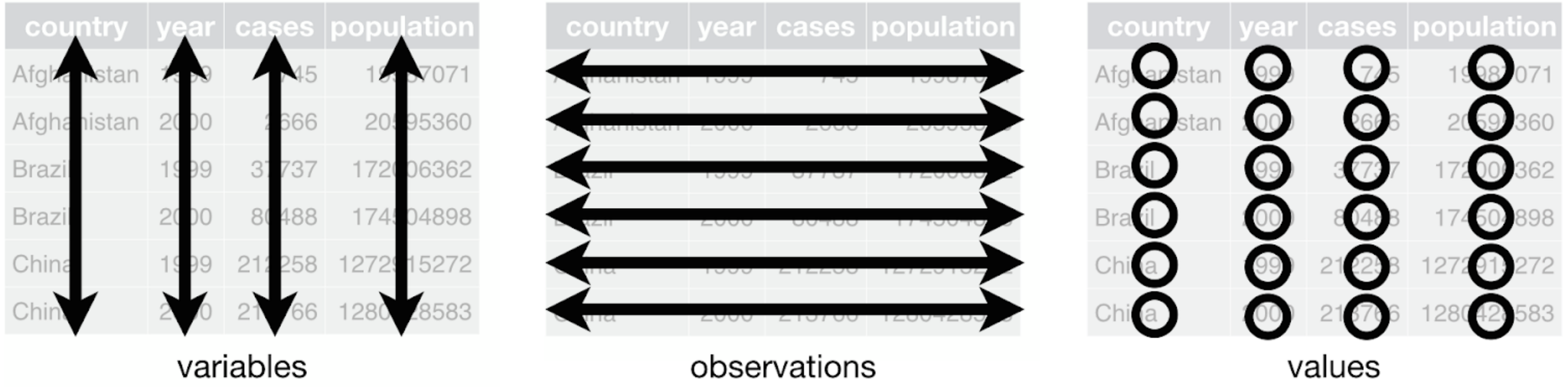
lau@ucsd.edu

OH: Wed 10-11a in SSRB 100

Sam OH moved to Mon 11-12pm, disc will be OH!

Tidy Data

- **Data table with a particularly useful format:**
  - **Every variable has its own column**
  - **Every observation has its own row**
  - **Every value has its own cell**



https://r4ds.had.co.nz/tidy-data.html

Tidy Data

Tidy

	country	year	cases	population
0	Afghanistan	1999	745	19987071
1	Afghanistan	2000	2666	20595360
2	Brazil	1999	37737	172006362
3	Brazil	2000	80488	174504898
4	China	1999	212258	1272915272
5	China	2000	213766	1280428583

Not tidy!

	country	1999	2000
0	Afghanistan	745	2666
1	Brazil	37737	80488
2	China	212258	213766

Tidy Data

Tidy

	country	year	cases	population
0	Afghanistan	1999	745	19987071
1	Afghanistan	2000	2666	20595360
2	Brazil	1999	37737	172006362
3	Brazil	2000	80488	174504898
4	China	1999	212258	1272915272
5	China	2000	213766	1280428583

Not tidy!

	country	year	type	count
0	Afghanistan	1999	cases	745
1	Afghanistan	1999	population	19987071
2	Afghanistan	2000	cases	2666
3	Afghanistan	2000	population	20595360
4	Brazil	1999	cases	37737
5	Brazil	1999	population	172006362

Use **pd.melt** to turn columns into rows and **pd.pivot_table** to turn rows into columns.

seaborn

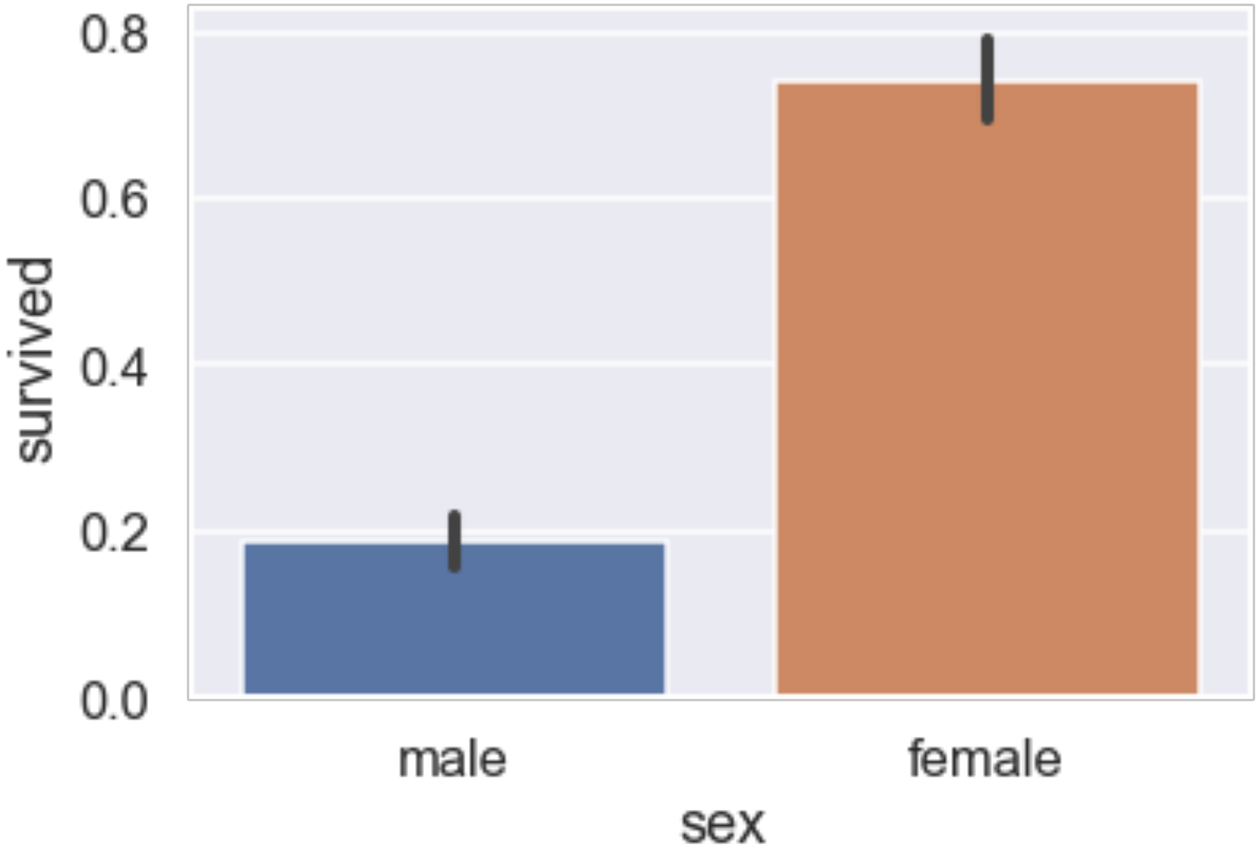
seaborn

- **Makes common statistical plots from tidy data.**
- **Typical usage:**

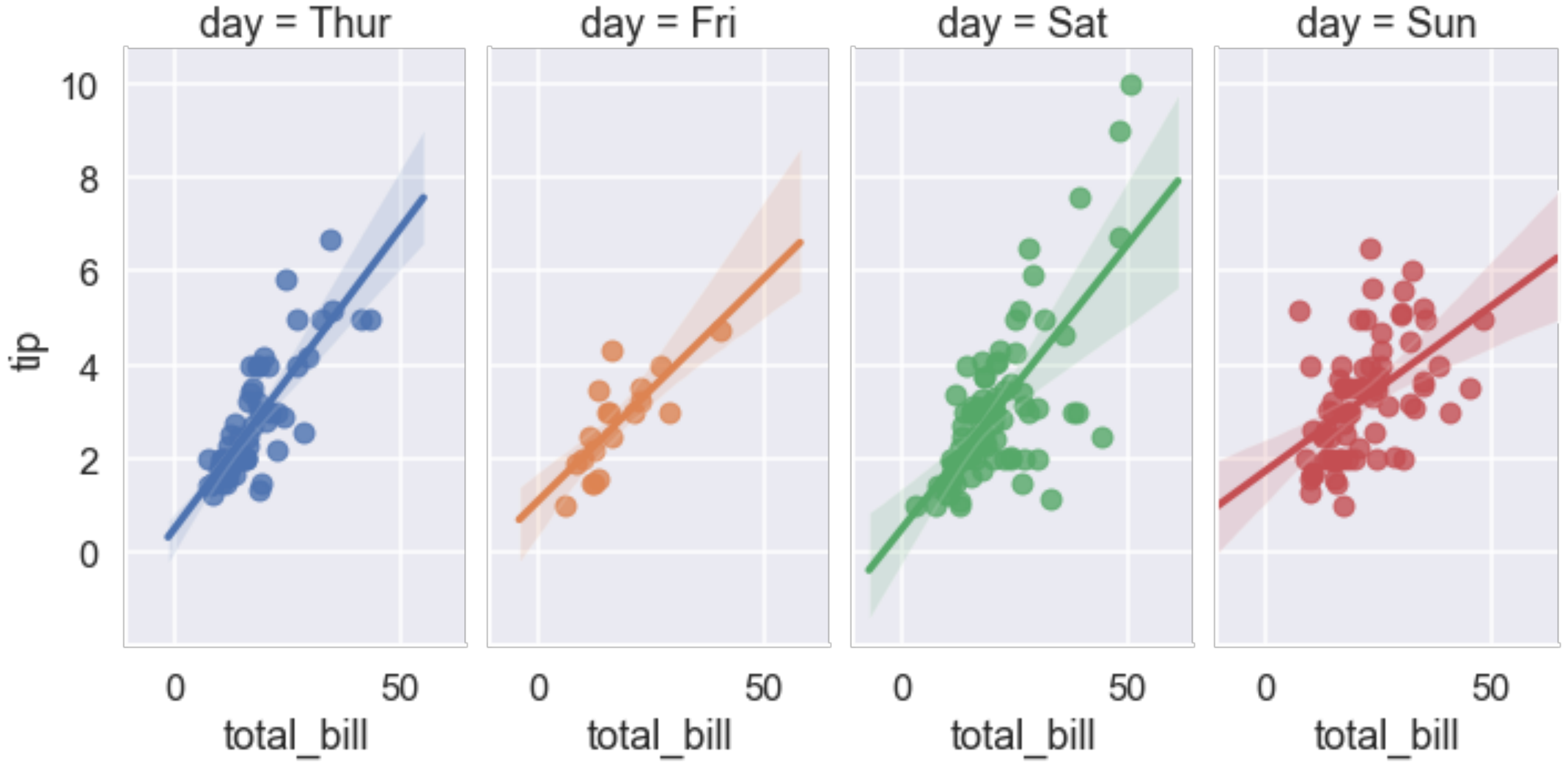
```
sns.someplot(x='...', y='...', data=...)
```

```
sns.barplot(x='sex', y='survived', data=ti)
```

	survived	class	sex	age	fare
0	0	Third	male	22.0	7.25
1	1	First	female	38.0	71.28
2	1	Third	female	26.0	7.92
...	...	...	...	...	...
888	0	Third	female	NaN	23.45
889	1	First	male	26.0	30.00
890	0	Third	male	32.0	7.75



```
sns.lmplot(x="total_bill", y="tip",
           col="day", hue="day", data=tips)
```



seaborn demo:

bit.ly/108-sam08

Work on A4!