

QSS20: Modern Statistical Computing

Unit 14: Work session

Goals for today

- ▶ **Recap of SQL**
- ▶ Upcoming deadlines
- ▶ Final project work session

Recap of SQL

What do you remember?

Recap of SQL

Tips:

- ▶ Use creds .yaml file to access connector—or copy into code (less secure)
- ▶ Subqueries in () are like functions as in `df.apply(func)`
- ▶ Use `count(*)` with `group by` to aggregate each slice (e.g., by race x crime)

Simple ex. w/ basic syntax:

```

1 select CASE_ID, AGE_AT_INCIDENT    —select cols
2 from caseinit                     —tablename
3 —table ops like joins or subqueries would go here
4 where AGE_AT_INCIDENT > 40        —condition(s)

```

Ex. creating new col w/ case/when:

```

1 select *                          —take all cols
2 CASE
3     WHEN OFFENSE_CATEGORY = UPDATED_OFFENSE_CATEGORY
4     THEN 'Same offense'
5     ELSE 'Diff offense'
6 END as charge_update
7 from caseinit

```

Where we are

- ▶ Recap of SQL
- ▶ **Upcoming deadlines**
- ▶ Final project work session

Upcoming deadlines

- ▶ Problem set five
 - ▶ **Due this Friday, 03/03**
 - ▶ Submit in usual way (assign Prof. & TAs a GitHub issue)
- ▶ Final project presentation
 - ▶ **Delivered in class this coming Monday, 03/06**
 - ▶ Aim for 8-12 mins; then 5-8 mins for Q&A
 - ▶ Slides due before class:
 - ▶ PDF → Canvas
 - ▶ **share w/ Prof. on Overleaf** (as with milestone 1)
- ▶ Final paper
 - ▶ **Due Monday, 03/13** (last day of scheduled finals)
 - ▶ How to submit:
 - ▶ PDF → Canvas
 - ▶ **share w/ Prof. on Overleaf**

More info on status-update presentations

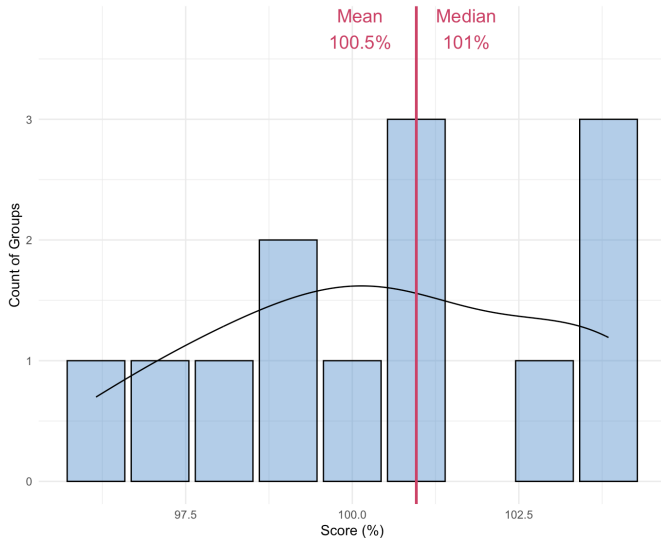
Expectations and what to expect:

- ▶ We will bring pizza! Requests?
- ▶ Please show as little code as possible (too dense)
- ▶ Please ask other groups questions; leave your slides alone in class
- ▶ QSS faculty guests may be here; can invite others (let Prof. know)

Order of group presentations:

1. **Numpy Narwhals** (SIP: SIRS)
2. **BACE** (SIP: SIRS)
3. **Lambdas** (SIP: SIRS)
4. **SeeSaw** (SIP: Medical training)
5. **Caspian** (SIP: Medical training)
6. **Will** (undergraduate religiosity)
7. **Dynamic Duo** (education finance)
8. **Prisoner's Dilemma** (felony sentencing)

Pset 4 grades: Strong (again)



Where we are

- ▶ Recap of SQL
- ▶ Upcoming deadlines
- ▶ **Final project work session**

Final project work session

Please return to class by 5:10 to debrief!

Things to work on:

- ▶ Slides for final presentation
- ▶ Analysis, codebase development, documentation
- ▶ Final paper

Places to work:

- ▶ This classroom
- ▶ Outside in hall
- ▶ Rockefeller Center atrium (up & downstairs—**please mind your volume**)

Feel free to consult with/ask questions of:

- ▶ Prof. Haber
- ▶ Ramsey
- ▶ Other groups/research teams

Last thing: Notecards

What's one thing you learned/made progress on today and one lingering challenge or question?