



INFORMATION ABOUT IN-CLASS ASSESSMENT 01

PSTAT 100: Data Science Concepts and Analysis
Spring 2024, with Ethan P. Marzban

LOGISTICAL INFORMATION ABOUT THE ICA

- The ICA will begin at 9:30 AM PDT in **EMBARCADERO HALL** (please note that this is **NOT** our usual lecture classroom).
 - The address for Embarcadero Hall is 935 Embarcadero del Norte, Isla Vista, CA 93117; the building is located in Isla Vista, right across from *Woodstocks* Pizza.
- The ICA will last for **60 minutes**, and will consist entirely of multiple choice questions.
- You are **allowed** the use of a **calculator**, and are **encouraged** to bring one as some questions may ask you to perform simple arithmetic calculations (e.g. finding the mean of a list of numbers).
 - Any type of calculator is allowed (including graphing calculators), so long as it cannot connect to the internet (so no using your phone as a calculator!)
 - Neither myself nor the proctors will have spare calculators - if you need a calculator, you can check one out from the library.
- You will be provided with the R cheatsheets that have been on the course computing server since the start of the quarter, along with some additional definitions, formulas, and help files on certain R functions.
 - You are **not** allowed the use of any additional notes; again, all necessary information will be provided to you with the ICA.
- Please bring your **access card** (or another valid form of identification) so we can verify your identity.
- If you have made arrangements for **DSP** accommodations (again, these would have had to have been submitted to and handled through the DSP office; we cannot honor any accommodation requests that don't come from DSP directly), please follow all information provided to you by DSP itself.
 - Students with DSP accommodations will not be taking the ICA in Embarcadero Hall, but in a location decided by DSP.
 - We cannot offer time extensions to anyone who takes the ICA in Embarcadero Hall, as we only have the room booked until 10:45 am.
- **Seating will be assigned**, and the seating chart will be posted to Canvas by 5pm on Wednesday April 24 (the evening before the ICA).
- **Nobody will be allowed to leave in the last 10 minutes of the ICA**, though if you finish earlier than 10 minutes from the end of the ICA you will be permitted to leave early.

FAQ (FREQUENTLY ASKED QUESTIONS)

- **How many questions will be on the ICA?**

At this point in time, I do not know exactly - the plan is to have somewhere between 20 and 40 questions (again, all multiple choice), but the exact number will depend on how involved the questions are. Furthermore, I will not be releasing the exact number of questions on the ICA before the ICA - everyone will find out at the same time, once the ICA begins!

- **What's the policy on calculators/notes?**

Please see the points in the above section - calculators are allowed (and graphing calculators are fine), but no additional aids (including notes) are allowed.

- **What's the bathroom policy during the ICA?**

If you end up needing to use the restroom during the ICA, please raise your hand and call a proctor over to your seat. The proctor will ask you to leave all of your things on your desk, and will hold onto your ID until the ICA ends.

- **Will a past or practice ICA be posted anywhere?**

Unfortunately I do not have any ICAs from previous quarters that I can post. I am, however, including a set of some practice questions in this document which you can use to help prepare you for the ICA.

- **How do you recommend I study for the ICA?**

This is a great question, with a potentially unsatisfactory answer - it really depends! Everyone has different study habits, and there isn't a single one-size-fits-all approach to studying for an assessment.

One piece of advice I can give is to go through some of the "Your Turn" exercises from the lecture slides, as these will be a good starting point for studying the "theoretical" or "conceptual" parts of the class. I'd also recommend going through the labs, homework, and mini project and focus on the "high-level picture" - rather than trying to figure out exactly what command does what, think more about which command to use where, and why we might want to use these commands (e.g. why was it, in Lab02, that we needed to melt and then pivot our dataframe?)

- **Will I be required to write code by hand?**

No; as stated above, the majority of questions will be multiple choice. I may, however, ask you to interpret the output of some code, or give you a series of code options and ask "which of the following will achieve ____?". The practice problems (which will be posted by Friday or Saturday, April 19 or 20) should hopefully clarify.

LIST OF TOPICS/TERMS

Please note: this is not meant to be a fully comprehensive list of topics. Just because a topic doesn't show up on this list, doesn't mean it isn't potentially testable on ICA01. (As a general rule-of-thumb: if a topic appeared in Lecture (including demos!), lab, homework, or on a mini-project, then it is fair game for the ICA.) On the other hand, just because a topic *does* appear in this document doesn't necessarily mean it *will* be tested on ICA01.

Fundamentals of Data:

- The Data Science Lifecycle
- Data semantics (observational units, observations, variables)
- Data structure (rows, columns, rectangular vs. nonrectangular representations of data)
- Tidy data (the three principles of data tidying; why we care about tidying data)
- Variable classification (numerical vs. categorical; discrete vs. continuous; ordinal vs. nominal)

Data Visualization:

- Basic types of graphs (barplot/bargraph, histogram, boxplot, scatterplot, line graph side-by-side boxplot, balloonplot)
 - Notions of trend and association (linear/nonlinear, positive/negative, etc.)
- Grammar of graphics (aesthetics, layers, geoms)
- Encoding information by modifying aesthetics
- Principles of color/color theory
 - RGB specification, hex color specification
 - CVD [deuteranomaly/deutanopia, protanomaly/protanopia, tritanomaly/tritanopia]
 - Color scales (qualitative, sequential, diverging)
 - Principles of good visualizations (from lecture 4)

Study Design and Sampling:

- Observational study vs. experiment
- Longitudinal study vs. cross-sectional study
- Sampling terms (total population, access frame, sampled population, sample)
- Census
- Probabilistic sampling techniques (simple random sampling, stratified random sampling, cluster sampling)
 - Inclusion probability
- Bias (selection bias, non-response bias, overcoverage, undercoverage)
- Scope of inference
- Causal relationships/effects, and confounding variables

Data Manipulation / R:

- Applying scale transformations to axes in a plot
- The pipe operator
- Data wrangling functions like `group_by()`, `summarise()`, `mutate()`, and `rename()`
- Melting dataframes
- Pivoting dataframes
- Merging/joining dataframes
 - Basic database concepts (primary key, compound key, foreign key relationship)
- Dataframes in R
 - Accessing columns using the `$` operator
 - Accessing elements using indexing/subsetting
- Any material from Lab00 (prerequisite knowledge)