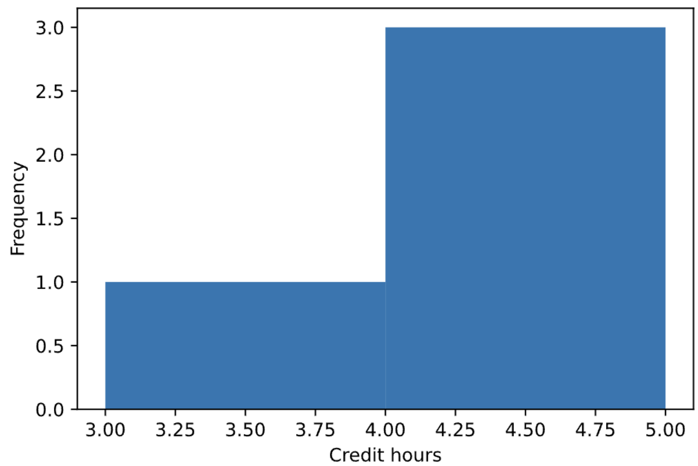
Exam 1

PUBPOL 2130 / INFO 3130
Moon Duchin, Spring 2025

(1) This is a screenshot from your Week1 notebook.

```
In [ ]: # This takes the credit hours from the course table above and plots them as a histogram.
        ax = df["credits"].plot.hist(bins=np.arange(3, 6))
        ax.set_xlabel("Credit hours")
        plt.show()
```



This is not a great data product and the next cell in the notebook corrects it. What is wrong with this histogram? There are many ways to correct it – draw a sketch of a better-looking histogram for this data and give some code that would create it. (Doesn't matter if it's the same fix that was in your notebook.)

The notebook fix was focused on the x-axis labels, which are ambiguous – which side of the 4.00 tick do the courses with four credit hours fall on?

Other reasonable student answers included simplifying the numbers on both axes so that they are only whole numbers, and adding a title. Some felt that there should be space between the bars so that it's clearer that the data only takes integer values.

The notebook fix was

```
ax = df["credits"].plot.hist(bins=np.arange(3, 6) - 0.5)
ax.set_xticks(range(3, 5))
```

But any working code related to that (or other fixes) was fine too! You were graded on showing some savvy in discussing a good visualization, and on showing some competence with Python to accomplish it.

(2) Here is the head of a geodataframe containing VAP (voting age population) data.

Out[5]:

	STATEFP20	COUNTYFP20	TRACTCE20	GEOID20	total_vap	geometry
0	12	063	211100	120632111003148	0	POLYGON ((-85.36774 30.57276, -85.36749 30.573...
1	08	005	006400	080050064002016	27	POLYGON ((-104.99384 39.6259, -104.99267 39.62...
2	12	086	011203	120860112033010	25	POLYGON ((-80.48766 25.47983, -80.48563 25.479...
3	08	039	961100	080399611002135	0	POLYGON ((-103.98078 38.86877, -103.98074 38.8...
4	08	029	964600	080299646004042	17	POLYGON ((-107.69036 38.81946, -107.68963 38.8...

It contains data from multiple states, including Colorado (state FIPS `"08"`). You want to know the total VAP in Arapahoe County (`08005`) plus Archuleta County (`08007`). Here are four attempts at writing code to do that. For each, say YES if it would work and NO if it would not. Briefly annotate to say what the code is doing or why it would fail. (If wrong, don't have to identify every mistake, just one for each.)

```
gdf[(gdf["STATEFP20"] == '08') & ((gdf["COUNTYFP20"] == '005') | (gdf["COUNTYFP20"] == '007'))]["total_vap"].sum()
```

YES. Here, & means AND and | means OR. This queries for rows that match the right state and either of the two counties, then specifies the total VAP column before summing. Incidentally, this is the format for these kinds of compound queries that was presented in your notebook.

```
gdf["total_vap"]["STATEFP20"=='08' & "COUNTYFP20" == '005' | "COUNTYFP20" == '007'].sum()
```

NO. Several fatal problems. One is that starting with `gdf["total_vap"]` restricts to just that column – now the state and county info is gone!

Another is that `"STATEFP20"=='08'` is an assertion of equality of string literals. It's not looking at the table! It will return FALSE.

```
gdf["STATEFP20" == '08' and "COUNTYFP20" == '005' or '007']["total_vap"].sum()
```

NO. Same problem with asserting equality rather than referencing the table, within the brackets. Also, `and` and `or` are not accepted in Pandas for those Boolean operations. Also, some parentheses are needed!

```
stategdf = gdf[gdf["STATEFP20"] == '08']
countygdf = stategdf[(stategdf["COUNTYFP20"] == '005') | (stategdf["COUNTYFP20"] == '007')]
countygdf["total_vap"].sum()
```

YES. This first constructs a new geodataframe consisting of just the rows with the right state. Then constructs another one consisting of the rows of that which match the right counties. Then it sums the VAP column.

For questions 3-7, your answer should be BRIEF—about three sentences! Succinct answers preferred.

(3) True/False and explain: The U.S. Census draws on race categories produced by a bureaucratic government office (the Office of Management and Budget). There are over 150 race/ethnicity categories that end up in the Decennial Census data.

Mostly true. It's true that the OMB makes the race categories. There have been six basic race choices since the 1970s (White, Black, Asian, AMIN, NHPI, and SomeOtherRace). There are just two ethnicity choices: Hispanic and Non-Hispanic.

Since the 1990s, you can choose multiple races. So you can choose yes/no on any race, but can't say no on all of them, making $2^6 - 1 = 63$ races once combinations are allowed. Then there are the two ethnicities, doubling the combinations so that there are 126 overall.

For this one, I wasn't looking for you to give the formula or know the exact number (and I don't care if you answered true or false)—I was looking for you to show some knowledge about how race/ethnicity work and what policies ended up producing so many of them.

(4) Short answer: Melissa Nobles writes, "Whereas in the United States, racial data are the raw material for civil rights legislation and policies, in Brazil, activists seek to recast the census as the destroyer, not the sustainer, of the racial democracy idea." What does she mean? How could racial data facilitate civil rights? By contrast, how might a census destroy the ideal of racial democracy?

There are two parts to this question.

First, how could race data provide raw material for civil rights legislation and policies—i.e., facilitate civil rights? One example is employment discrimination law, which allows lawsuits when companies systematically discriminate on the basis of race. Another example is the Voting Rights Act, which (sometimes) allows plaintiffs to push for districting plans that give members of minority groups an opportunity to elect candidates of choice.

Second, how might a census destroy the ideal of racial democracy, according to Brazilian activists? Nobles is referring to a *universalist* ideal where everyone is mixed in a multi-racial Brazilian polity, and identity mobility is possible. The fear she is citing is that having categories for race will invite Brazilians to think of themselves as different, rather than the same.

(5) True/False and explain: The Mason/Dixon line was a court-ordered solution to an eminent domain dispute. Mason was a business owner and Dixon was a zoning official. The line was hotly contested at the time but did not end up being important.

False. The Mason/Dixon line was court ordered, but it was a solution to a border dispute primarily between Maryland and Pennsylvania before the Revolutionary War. The reason for that dispute was that King Charles had made sloppy/imprecise promises to different parties, putting a small but key piece of land in contention. Mason and Dixon were land surveyors who were brought in by the court; they used physical markers to stake out a more precise boundary.

Later, it became famous as a symbolic dividing line between free states and slave states.

(6) Short answer: In class, there was an example of a page from a government-issued document describing "low-grade imbecile" as a category. Ian Hacking's ten elements in his making up / looping framework are: Count, Quantify, Create Norms, Correlate, Medicalize, Biologize, Normalize, Bureaucratize, Reclaim-Our-Identity. Not all ten apply to the "imbecile," but discuss at least two of these engines at work in this case. (And even if you don't remember this particular example, discuss briefly how and why a government might identify "imbeciles.")

I think the three that are most clearly related to the episode as we discussed it are Quantify, Create Norms, and Bureaucratize – though some made the case for other "engines."

Quantify: the category was defined with a metric (IQ) and a threshold.

Create Norms: by specifying ranges for "mental defectives," the publication implicitly defines the normal range. (For instance, the upper end of "moron" is at IQ 70, and 100 is supposed to be average.)

Bureaucratize: this enables public policy triggered by the categorization. In this case, employment ("simple menial work") and reproductive rights (through eugenics) were both referenced.

(7) Making maps has always been a collaboration between people and technology for measuring and recording data. Briefly describe three examples of human involvement or technological involvement—or a human/technology collaboration—that were significant steps or tools in the process of mapmaking at some point in history. (You can draw on lectures, the Scott reading, the Monmonier reading, etc. Show off some knowledge from this class!)

Some examples to get you thinking: Tanzania (place), the Pope (person), a theodolite (thing), metes and bounds (notation).

I'll just quote a few strong student answers! (lightly edited)

- Another advancement was in how borders are described and defined. They used to be described using metes and bounds (English descriptions of the border based on landmarks, requiring local knowledge.) Now, GIS systems use vector-formatted data and units like census blocks to define which places are within boundaries.
- Shapefiles—technology. A form of data storage used to record information about a place or places based on attributes attached to polygonal units. There are flaws with this, including limited attributes or a lack of differentiation between 0 and empty.
- Gauss was an important figure in the history of mapmaking. He came up with an abstract definition of curvature that related to concrete measurements. This contributed to the theory of mapping data from curved to flat surfaces (or vice versa), where you can prove that angles and other attributes can not be simultaneously preserved. This led to modern map projections seeking compromises.
- The choropleth is a common modern cartographic method used for recording data in a map through a colorful visualization. One can identity a region's attributes, based on shading and color. Making choropleths in GIS requires human judgment about which attributes, colors, and shading gives the most insight, as well as technology to join the needed attributes to the units being pictured.
- Tanzanian villagization, Soviet collectivization, and China's Great Leap Forward are all historic examples mentioned by Scott where a combination of high modernist ideology, authoritarianism, and a weakened civil society contributed to disaster. All of these were driven by simplification/standardization of society, which was rooted in a technological/institutional process of mapping land, agriculture, and property to facilitate taxation and other state operations.
- Sonar and other underwater technology allowed us to collect more data from the sea and to explore the sea bed (e.g., for extracting resources). This made deciding sea boundaries between countries much more important.