DATA 6
Fall 2025

Lisa Yan
Quiz 1

Your name: _____

Your student ID: _____

Your Berkeley email: _____

Your room location: _____

Student ID of the person to your left: _____

Student ID of the person to your right: _____

You have 50 minutes. There are 3 questions of varying credit. (36 points total)

Question:	HC	1	2	3	Total
Points:	1	7	28	0	36

For questions with **circular bubbles**, you may select only one choice.

- ○ Unselected option (Completely unfilled)
- ◌ Don't do this (it will be graded as incorrect)
- ● Only one selected option (completely filled)

For questions with **square checkboxes**, you may select one or more choices.

- ■ You can select
- ■ multiple squares
- ☑ (Don't do this)

Anything written outside the answer boxes or ~~crossed out~~ will not be graded. If you write multiple answers, your answer is ambiguous, or the bubble/checkbox is not entirely filled in, we will grade the worst interpretation. For coding questions with blanks, you may write at most one statement per blank and you may not use more blanks than provided.

As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. I will follow the rules of this exam.

Honor Code (HC): I have read and agree to the honor code above.

(1 point) Sign your name: _____

This content is protected and may not be shared, uploaded, or distributed.

SID: _____

Q1 *What Would Python Do? (WWPD)* **(7 points)**

Q1.1 (1 point) [True/False] In Data 6, a "variable" and a "name" refer to the same concept.

◯ `True` ◯ `False`

Q1.2 (1 point) [True/False] In Data 6, a "variable type" and a "data type" refer to the same concept.

◯ `True` ◯ `False`

Q1.3 (1 point) Consider the below code:

```
1 x = 0
2 y = x + 1
3 x + 1
4 x = x + 1
```

After this code is run, what is the value of `x`?

◯ `0` ◯ `1` ◯ `2` ◯ `3`

Q1.4 (2 points) Recall that `print` displays arguments separated by spaces (`' '`). Further, recall that `'\n'` (or `"\n"`) inserts a newline (i.e., line break) in a string.

When the below code is run, what is printed?

```
s = "DATA" + "\n" + "6"
print(s, s)
```

◯
```
DATA 6
DATA 6
```

◯
```
DATA
6 DATA
6
```

◯ None of these

◯
```
DATA
6
DATA
6
```

◯
```
DATA
6DATA
6
```

Recall that functions can have names, too. For Q1.5-Q1.6, consider the below code:

```
1 len = sum
2 arr = make_array(1, 2, 3, 4)
3 arr2 = make_array(3, 1)
4 print(min(len(arr), len(arr2)))
```

Hint: In line 1, `sum` is a built-in Python function; see the cheatsheet.

Q1.5 (1 point) In line 4, what does `len(arr2)` evaluate to?

◯ `2` ◯ `4` ◯ `10` ◯ None of these

Q1.6 (1 point) When the code is run, what is printed?

◯ `2` ◯ `4` ◯ `10` ◯ None of these

This content is protected and may not be shared, uploaded, or distributed.

SID: _____

Q2 *BRFSS Table* **(28 points)**

Consider the Behavioral Risk Factor Surveillance System (BRFSS), a health survey fielded by the Centers for Disease Control and Prevention (CDC). We have created `brfss`, a table constructed from a subset of the variables computed from BRFSS survey responses (Table 1). Each row in `brfss` represents one individual's responses to the survey.

ID	State	Flu Shot?	Education Level	Height (in)
34428	South Dakota	No	Graduated High School	64
59206	Delaware	Don't Know/Blank	Did not graduate High School	65
66594	Florida	Yes	Attended College	66
67207	Maryland	Yes	Graduated High School	68
88712	Michigan	Yes	Graduated College	64

Table 1: The first 5 rows of the `brfss` table (455,122 rows total).

Variable descriptions:
- `ID`: A 5-digit number uniquely identifying the respondent.
- `State`: State of residence.
- `Flu Shot?`: Has taken the annual flu vaccine.
- `Education level`: Level of education completed.
- `Height (in)`: Reported height in inches (integer values).

Q2.1 (1 point) What is the variable type of `ID`?

◯ Discrete Numerical ◯ Ordinal Categorical

◯ Continuous Numerical ◯ Nominal Categorical

Q2.2 (1 point) What is the variable type of `State`?

◯ Discrete Numerical ◯ Ordinal Categorical

◯ Continuous Numerical ◯ Nominal Categorical

Q2.3 (1 point) What is the variable type of `Flu Shot?`?

◯ Discrete Numerical ◯ Ordinal Categorical

◯ Continuous Numerical ◯ Nominal Categorical

Q2.4 (1 point) What is the variable type of `Education Level`?

◯ Discrete Numerical ◯ Ordinal Categorical

◯ Continuous Numerical ◯ Nominal Categorical

Q2.5 (1 point) What is the variable type of `Height (in)`?

◯ Discrete Numerical ◯ Ordinal Categorical

◯ Continuous Numerical ◯ Nominal Categorical

This content is protected and may not be shared, uploaded, or distributed.

Consider the below definition of `height`, to be used for the remainder of the question.

```
height = brfss.column("Height (in)")
```

Q2.6 (1 point) What is the data type of `height`, i.e., what does `type(height)` evaluate to?

- ◯ `int`
- ◯ `numpy.ndarray`
- ◯ None of these
- ◯ `str`
- ◯ `datascience.tables.Table`

Q2.7 (5 points) Which of the below lines of code will assign `max_ht` to the maximum height of all respondents? Select all that apply.

- ☐ `max_ht = max(height)`
- ☐ `max_ht = brfss.sort("Height (in)").column("Height (in)").item(0)`
- ☐ `max_ht = brfss.sort("Height (in)").column("Height (in)").item(-1)`
- ☐ `max_ht = brfss.sort("Height (in)").height.item(-1)`
- ☐ `max_ht = max(brfss.select("Height (in)"))`
- ◯ None of the above

Q2.8 - Q2.9 (2 points) Assume that `max_ht` is correctly assigned in the previous part (i.e., `max_ht` is the maximum height of all respondents). Use `max_ht` to fill in the following line of code to compute the *number* of respondents who reported the maximum height.

```
brfss.where("Height (in)", _____Q2.8_____).____Q2.9____
```

Q2.10 - Q2.13 (6 points) Convert each height in `height` to its equivalent value in feet and inches (e.g., 1 foot is 12 inches, so 64 inches converted is 5 foot 4 inches).

Complete the code below by using `height` to create Table 2, a table of converted heights with `ft` and `in` columns.

Height	ft	in
64	5	4
65	5	5
66	5	6
68	5	8
64	5	4

Table 2: The first 5 rows of the table output when your code is run (455,122 rows total).

```
Table().with_columns("Height", height,
                     _____Q2.10_____, _____Q2.11_____,
                     _____Q2.12_____, _____Q2.13_____)
```


This content is protected and may not be shared, uploaded, or distributed.

Consider the `ed_table` in Table 3 (left), which aggregates and *counts* the number of respondents by their education level completed. We would like to modify `ed_table` to Table 4 (right), which reports the *percentage of respondents* per education level. Assume the possible education levels are the four values shown.

Education Level	# Respondents
Did not graduate High School	26011
Graduated High School	108990
Attended College	120252
Graduated College	187496

Table 3: Original `ed_table`.

Education Level	% Respondents
Did not graduate High School	5.87489
Graduated High School	24.6167
Attended College	27.1603
Graduated College	42.3481

Table 4: Modified `ed_table`.

Below, `ed_counts` is an array of the counts of respondents per education level.

```
ed_counts = make_array(26011, 108990, 120252, 187496)
```

Q2.14 (1 point) Consider `108990`, the count of respondents who reported "Graduated High School." What is this element's index in `ed_counts`?

○ 0 ○ 1 ○ 2 ○ 3 ○ 4

Q2.15 (3 points) Fill in the following line of code to assign `ed_percents` to an array of the *percentages* of respondents per education level.

Your answer should use `ed_counts` but should not round values.

```
ed_percents = _____
                Q2.15
```

Q2.16 - Q2.17 (5 points) Assume that `ed_percents` is correctly assigned in the previous part (i.e., `ed_percents` is an array of the percentages of respondents per education level) and that `ed_table` is Table 3 (left, the original table). Use `ed_percents` and `ed_table` to write code that modifies `ed_table` to Table 4 (right).

Note: Depending on your approach, you may not need both lines below. If you do not use the second line, cross it out.

```
ed_table = _____
             Q2.16

_____
Q2.17
```


This content is protected and may not be shared, uploaded, or distributed.

SID: _____

Q3 *Just for fun!* **(0 points)**

Q3.1 Draw something fun, or write a message for the staff! Or leave this blank!

This content is protected and may not be shared, uploaded, or distributed.