Name:

SA433 · Data Wrangling and Visualization

Fall 2024 · Uhan

Quiz 2 – 9/18/2024

Score	
	/ 10

Instructions. You have 15 minutes to complete this quiz. You may <u>not</u> use any outside materials. This quiz is closed computer, closed book, closed notes, and closed internet. No collaboration allowed.

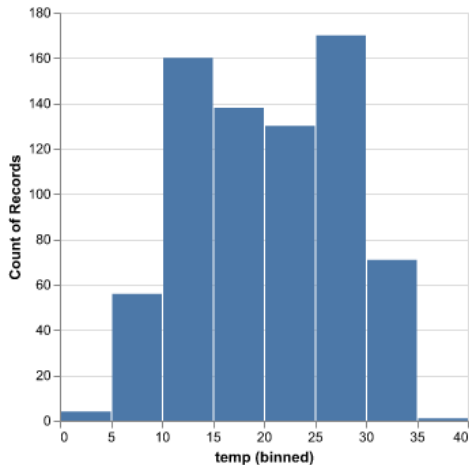The problems in this quiz are based on data from a bike sharing company. The data contains the following variables:

Variable	Description
date	The date in the format YYYY-MM-DD
dayofweek	The day of the week (0=Sunday, 1=Monday, etc)
workday	0 if date is not a workday, 1 if date is a workday
temp	Daily high temperature in degrees celcius
humidity	Percent humidity
casual	The number of rides by casual (non-registered) riders
registered	The number of rides by registered riders

The dataset was read into a Pandas DataFrame called `df`. Here are the first 5 rows of the DataFrame:

```
     date        dayofweek  workday       temp  humidity  casual  registered
0  2018-01-01            6        0  14.110847   80.5833     331         654
1  2018-01-02            0        0  14.902598   69.6087     131         670
2  2018-01-03            1        1   8.050924   43.7273     120        1229
3  2018-01-04            2        1   8.200000   59.0435     108        1454
4  2018-01-05            3        1   9.305237   43.6957      82        1518
```

Problem 1. Which code generates a histogram (as displayed below) to show the distribution of daily high temperatures represented in the data?
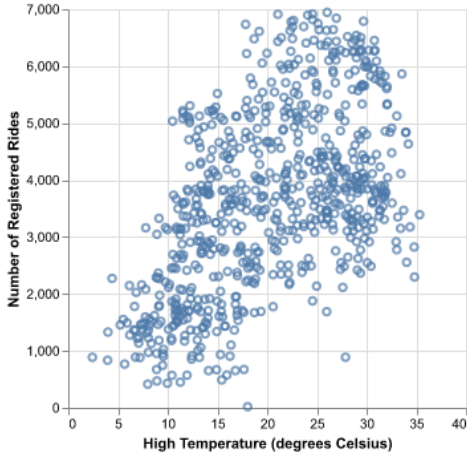

A.
```
alt.Chart(df).mark_bar().encode(
    alt.Y('count(temp):Q').bar()
)
```

B.
```
alt.Chart(df).mark_bar().encode(
    alt.X('temp:Q').bin(),
    alt.Y('count()')
)
```

C.
```
alt.Chart(df).mark_hist().encode(
    alt.X('temp:Q').bin(True),
    alt.Y('number()')
)
```

D.
```
alt.Chart(df).mark_hist().encode(
    alt.X('temp:Q'),
    alt.Y('count(registered):Q')
)
```

Your answer:

Problem 2. Which code generates the following scatterplot of high temperature versus number of rides with axis titles?
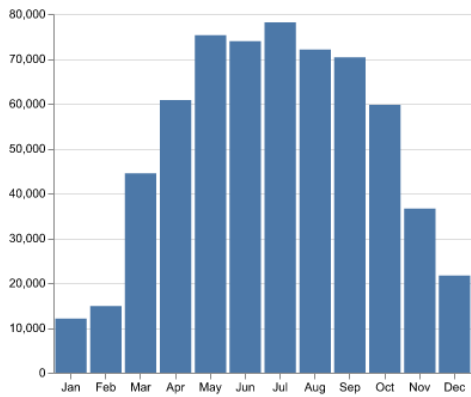
A.
```
alt.Chart(df).mark_point().encode(
    alt.X('temp:Q').title='High Temperature (degrees Celsius)'
    alt.Y('registered:Q').title='Number of Registered Rides'
)
```

B.
```
alt.Chart(df).mark_point().encode(
    alt.X('temp:Q'),
    alt.Y('registered:Q')
).configure_title(
    title.X('High Temperature (degrees Celsius)'),
    title.Y('Number of Registered Rides')
)
```

C.
```
alt.Chart(df).mark_point().encode(
    alt.X('temp:Q'),
    alt.Title('High Temperature (degrees Celsius)'),
    alt.Y('registered:Q'),
    alt.Title('Number of Registered Rides')
)
```

D.
```
alt.Chart(df).mark_point().encode(
    alt.X('temp:Q').title('High Temperature (degrees Celsius)'),
    alt.Y('registered:Q').title('Number of Registered Rides')
)
```

Problem 3. Which code generates the following bar graph, which presents the total number of casual rides in each month? (Note that the axis titles have been cropped out.)
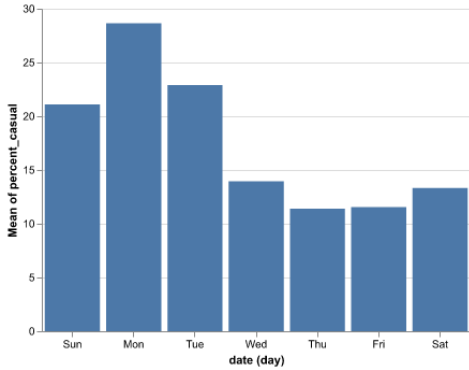
A.
```
alt.Chart(df).mark_bar().encode(
    alt.X('month(date):O'),
    alt.Y('sum(casual)')
)
```

B.
```
alt.Chart(df).mark_bar().encode(
    alt.X('month:Q'),
    alt.Y('casual:Q')
)
```

C.
```
alt.Chart(df).mark_bar().encode(
    alt.X('casual:Q').bin(type='month'),
    alt.Y('count()')
)
```

D.
```
alt.Chart(df).mark_bar().encode(
    alt.X('month(date):O'),
    alt.Y('casual:Q').sum()
)
```

Problem 4. Which code makes the bar graph below which displays the percentage of casual riders on each day of the week?
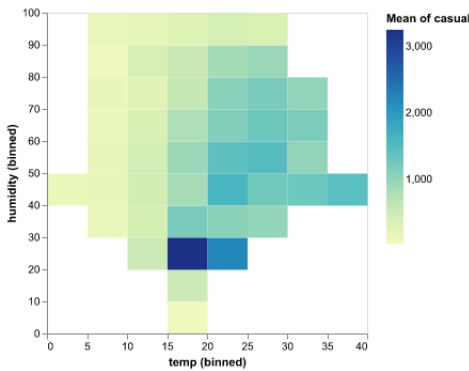
A.
```
alt.Chart(df).mark_bar().transform_filter(
    'percent_casual == 100 * datum.(casual/(casual + registered))'
).encode(
    alt.X('dayofweek:O'),
    alt.Y('mean(percent_casual):Q')
)
```

B.
```
alt.Chart(df).mark_bar().transform_calculate(
    percent_casual = '100 * datum.casual/(datum.casual + datum.registered)'
).encode(
    alt.X('dayofweek:O'),
    alt.Y('mean(percent_casual):Q')
)
```

C.
```
alt.Chart(df).mark_bar().make_variable(
    percent_casual = '100 * casual/(casual + registered)'
).encode(
    alt.X('dayofweek:O'),
    alt.Y('percent_casual:Q')
)
```

D.
```
alt.Chart(df).mark_bar().encode(
    alt.X('dayofweek:O'),
    alt.Y('mean.percent.casual:Q')
)
```

Problem 5. Which code generates the following two-dimensional heatmap visualizing the relationship between weather on the number of casual riders? Note that the x-axis represents temparature, the y-axis represents the humidity, and the color indicates the average number of casual riders.
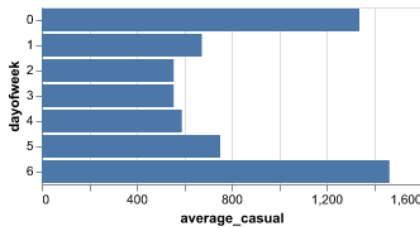
A.
```
alt.Chart(df).mark_histogram().encode(
    alt.X('temp:Q').bin(),
    alt.Y('humidity:Q').bin()
)
```

B.
```
alt.Chart(df).mark_square().encode(
    alt.X('temp:O'),
    alt.Y('humidity:O'),
    alt.Color('count()')
)
```

C.
```
alt.Chart(df).mark_rect().encode(
    alt.X('temp:Q').bin(),
    alt.Y('humidity:Q').bin(),
    alt.Color('mean(casual)')
)
```

D.
```
alt.Chart(df).mark_area().encode(
    alt.X('temp:Q').bin(),
    alt.Y('humidity:Q').bin()
).properties
    alt.Color('mean(casual)')
)
```

Problem 6. Which code snippet, when placed in the correct position, will modify an Altair chart of the bike sharing data so that it only displays observations corresponding to workdays?

A. `.transform_delete('workday == 1')`

B. `.transform_filter('workday == 0')`

C. `.transform_delete(datum.workday = 0)`

D. `.transform_filter('datum.workday == 1')`

Problem 7. The following code, which creates a bar chart showing the average number of casual riders per day of the week, is missing a function name:
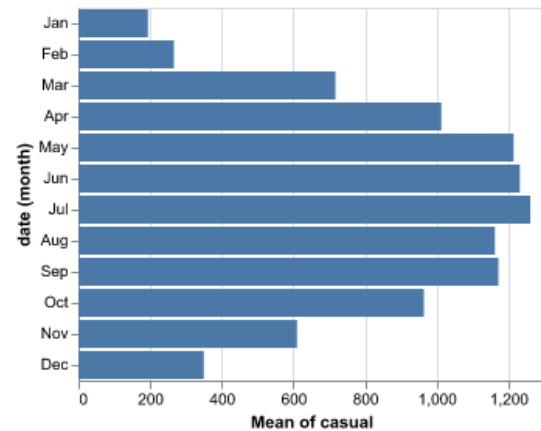
Your answer:


```
alt.Chart(df).A(
    groupby = ['dayofweek'],
    average_casual = 'mean(casual)'
).mark_bar().encode(
    alt.Y('dayofweek:O'),
    alt.X('average_casual:Q')
)
```

What is missing at "A"?

A. `transform_filter`

B. `transform_groupby`

C. `transform_aggregate`

D. `transform_calculate`

Scenario for Problems 8-10. Suppose we have written code to generate a bar graph that shows the average number of casual rides per day during each month. There are letters to mark various positions in the code. (No code is currently written at the marked positions.)


```
alt.Chart(df).mark_bar().A.encode(
    alt.Y('month(date):O').B,
    alt.X('mean(casual):Q').C,
    D
).properties(
    E
).configure_title(
    F
)
```

Where should we add code if we want to ...

Problem 8. ... add the chart title, "Average Daily Casual Rides by Month"?

Your answer:

Problem 9. ... change the font size of the chart title added in problem 8?

Your answer:

Problem 10. ... order the months by the mean number of casual rides?

Your answer: