Name: Solutions

SA433 · Data Wrangling and Visualization Fall 2024 · Uhan

# Exam 1 – Part 2 – 10/2/2024

| Score |
| --- |
| / 30 |

## Instructions

- This part is worth 30 points total. The exam (all three parts) is worth 100 points total.
- You have 50 minutes to complete Parts 1 and 2 of the exam.
- For Parts 1 and 2 of the exam, you may not use any outside assistance. These parts of the exam are closed book, closed notes, and closed internet.
- **No collaboration allowed.** All work must be your own.
- You must turn in Part 1 before beginning Part 2.
- **Do not discuss the contents of this exam with any midshipmen until it is returned to you.**

## Background

In this part of the exam, you'll use a dataset on carbon dioxide and greenhouse gas emissions, based on data from Our World in Data.

The dataset contains the following variables/columns for a subset of countries around the world, from 2000 to 2016:

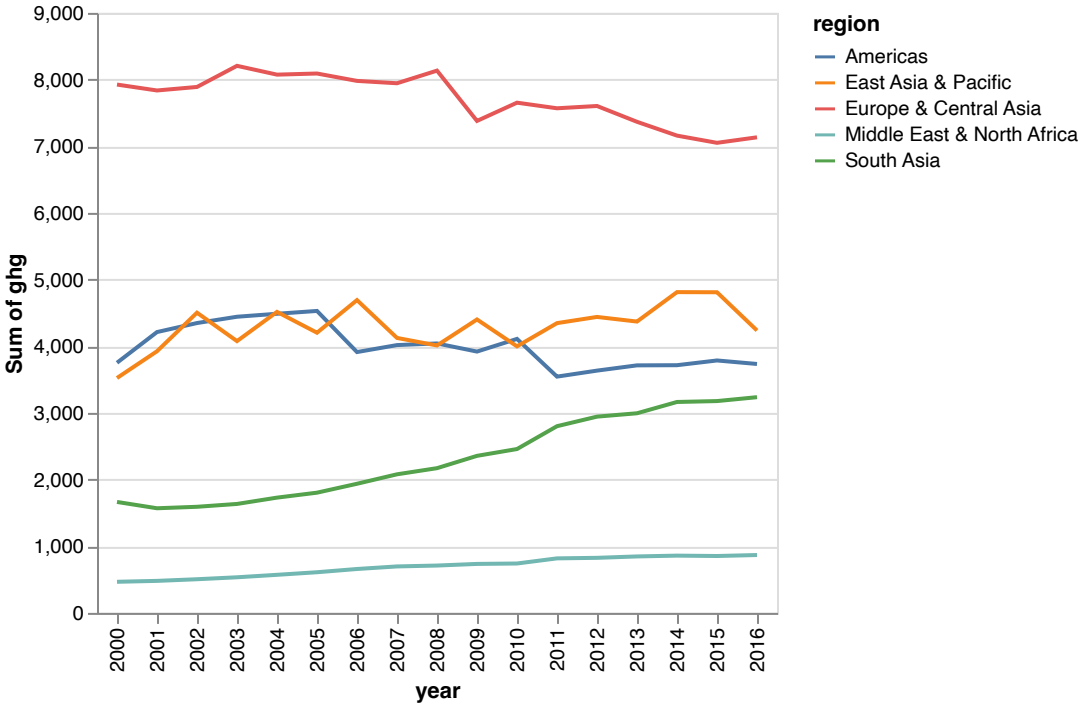
| Column | Description |
| --- | --- |
| `year` | Year of observation |
| `region` | Region of the world |
| `country` | Country |
| `co2` | Annual production-based CO2 emissions (million tonnes) |
| `ghg` | Annual greenhouse gas emissions (million tonnes of CO2 equivalents) |
| `population` | Total population of country |
| `gdp` | Total real GDP, inflation-adjusted |

The DataFrame is saved to the variable `df`. This is the result of `df.head()`:

| | year | region | country | co2 | ghg | population | gdp |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 2000 | Americas | Argentina | 141.717 | 366.34 | 36871000.0 | 5.570000e+11 |
| **1** | 2001 | Americas | Argentina | 133.311 | 383.27 | 37276000.0 | 5.250000e+11 |
| **2** | 2002 | Americas | Argentina | 124.382 | 386.17 | 37682000.0 | 4.530000e+11 |
| **3** | 2003 | Americas | Argentina | 134.621 | 408.49 | 38088000.0 | 4.870000e+11 |
| **4** | 2004 | Americas | Argentina | 157.034 | 436.71 | 38492000.0 | 5.320000e+11 |

**Problem 1.** The chart below shows the annual total green house gas emissions for each region. Your answer:

B




Which of the following code snippets produced this chart?

A.
```
alt.Chart(df).mark_line(colorby='region').encode(
    alt.X('year:O'),
    alt.Y('sum(ghg):Q')
)
```

B.
```
alt.Chart(df).mark_line().encode(
    alt.X('year:O'),
    alt.Y('sum(ghg):Q'),
    alt.Color('region:N')
)
```

C.
```
alt.Chart(df).mark_line().encode(
    alt.X('year:O'),
    alt.Y('sum(ghg):Q')
).properties(
    color='various'
)
```

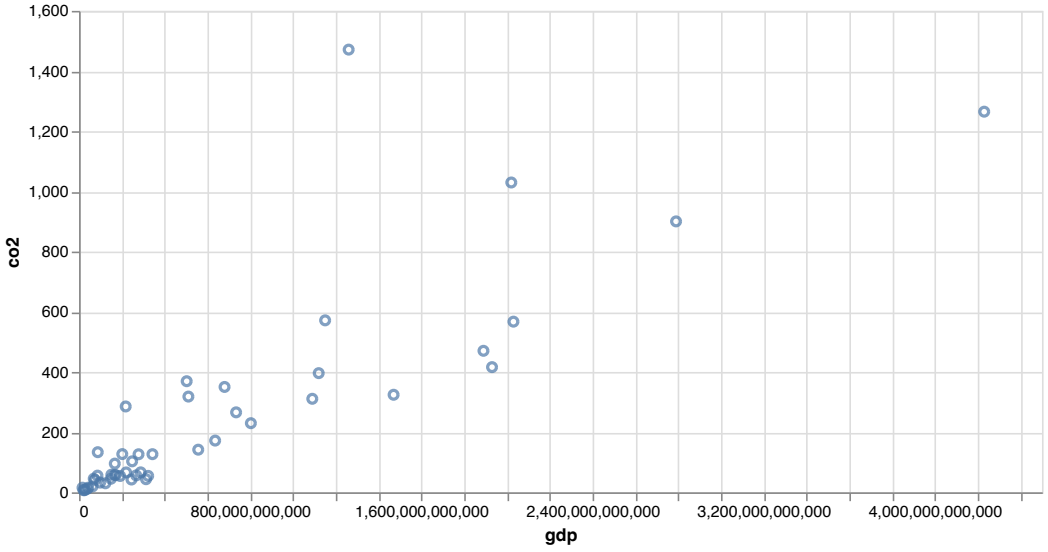
D.
```
alt.Chart(df).mark_line().encode(
    alt.X('year:O'),
    alt.Y('sum(ghg):Q'),
    alt.Color(groupby='region:N')
)
```

Name:

Your answer:

C

**Problem 2.** The following chart shows the inflation-adjusted GDP versus the production-based CO2 emissions for the countries in the data set.




This is the code that was used to make the chart. Note that there are letters (`#A`, `#B`, etc.) added to mark possible positions for additional code.

```
alt.Chart(df).transform_filter(
    'datum.year==2000',
    #A
).mark_point(
    #B
).encode(
    alt.X('gdp:Q'),
    alt.Y('co2:Q'),
    #C
).properties(
    width = 600,
    #D
)
```

Suppose you want to change the size of the points according to the population. Where would you add the appropriate code to make this change?

A. `#A`
B. `#B`
C. `#C`
D. `#D`
E. None of these.

**Problem 3.** Consider the chart and code in Problem 2. Suppose you want to change the size of all of the points uniformly. Where would you add the appropriate code to make this change?

Your answer:

B

A. `#A`
B. `#B`
C. `#C`
D. `#D`
E. None of these.

**Problem 4.** Consider the chart and code in Problem 2. Suppose you want to add an overall title to the chart. Where would you add the appropriate code to make this change?

Your answer:

D

A. `#A`
B. `#B`
C. `#C`
D. `#D`
E. None of these.

**Problem 5.** Consider the chart and code in Problem 2. Suppose you want to change the scale of the x-axis to spread out the points that are clustered together. You can do this by adding the code `.scale(type='sqrt')`. Where should this code be added?

Your answer:

E

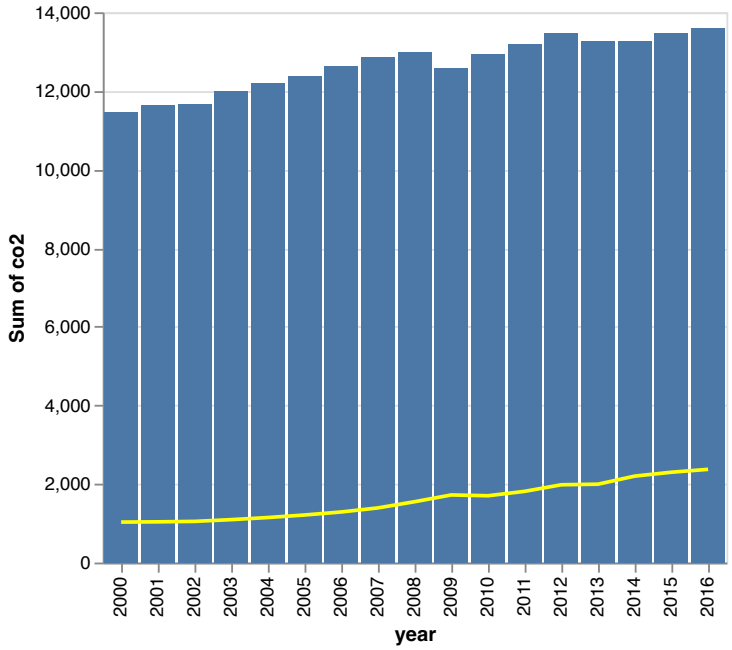
A. `#A`
B. `#B`
C. `#C`
D. `#D`
E. None of these.

Name:

**Problem 6.** The following chart shows the production-based CO2 emissions of countries in the South Asia region (the yellow line) as compared to the total emissions of all countries in the dataset (blue bars).

Your answer:

C




Consider the code:

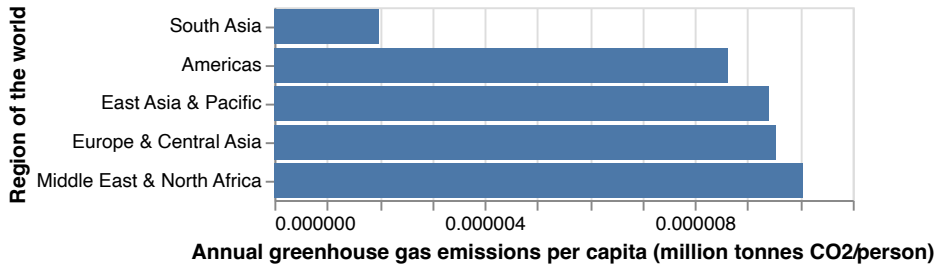
```
all_co2 = alt.Chart(df).mark_bar().encode(
    alt.X('year:O'),
    alt.Y('sum(co2):Q')
)

southasia_co2 = alt.Chart(df).transform_filter(
    'datum.region=="South Asia"'
).mark_line(color='yellow').encode(
    alt.X('year:O'),
    alt.Y('sum(co2)')
)
```

Combined with the code above, which of the following options produced the provided chart?

A. `all_co2 & southasia_co2`
B. `southasia_co2 | all_co2`
C. `all_co2 + southasia_co2`
D. `all_co2 - southasia_co2`

**Problem 7.** The following chart shows the annual greenhouse gas emissions *per capita* for each region in the year 2010.

Your answer: B




The chart was produced by the following code, except some of the code was replaced by `METHOD_ONE`, `METHOD_TWO`, and `SORT_CODE`:

```
alt.Chart(df).transform_filter(
    'datum.year == 2010'
).METHOD_ONE(
    groupby=['region'],
    region_ghg='sum(ghg):Q',
    region_pop='sum(population):Q',
).METHOD_TWO(
    region_ghg_per_capita = 'datum.region_ghg / datum.region_pop'
).mark_bar().encode(
    alt.X('region_ghg_per_capita:Q')
        .title('Annual greenhouse gas emissions per capita (million tonnes CO2/person)'),
    alt.Y('region:N')
        .title('Region of the world')
        .sort(SORT_CODE),
)
```

What code was replaced by `METHOD_ONE`?

A. `transform_calculate`
B. `transform_aggregate`
C. `transform_filter`
D. `transform_group`

**Problem 8.** Consider the chart and code in Problem 7. What code was replaced by `METHOD_TWO`?

Your answer: A

A. `transform_calculate`
B. `transform_aggregate`
C. `transform_filter`
D. `transform_group`

**Problem 9.** Consider the chart and code in Problem 7. What code was replaced by `SORT_CODE`?

Your answer: D

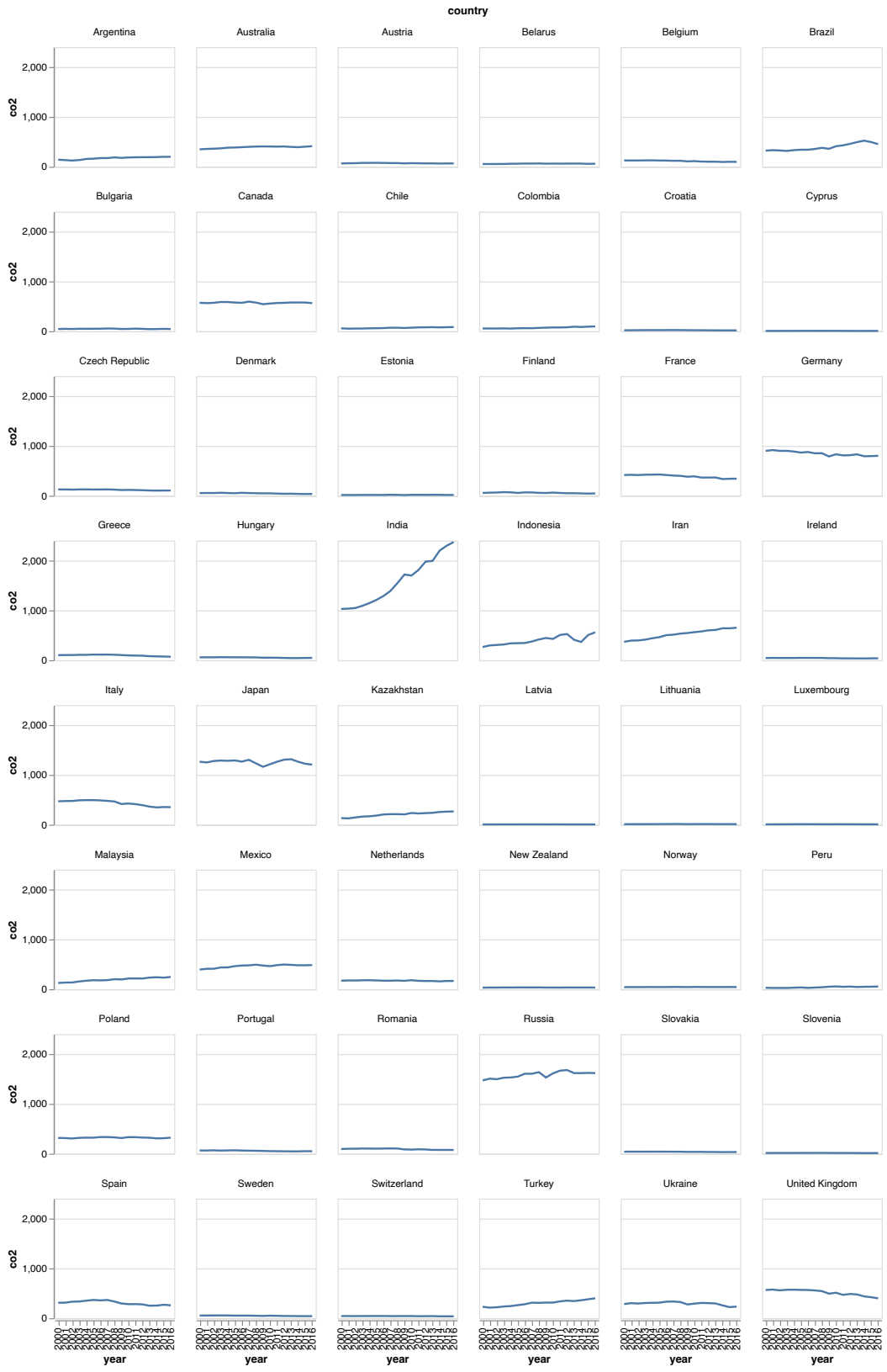
A. `{'field': 'x', 'order': 'descending'}`
B. `{'encoding': 'x', 'order': 'descending'}`
C. `{'field': 'x', 'order': 'ascending'}`
D. `{'encoding': 'x', 'order': 'ascending'}`

Name:

**Problem 10.** The chart below shows the CO2 emissions over time for each country in the dataset.

Your answer:

A




(cont.)

Which of the following code snippets produces this chart?

A.
```
base = alt.Chart(df).mark_line().encode(
    alt.X('year:O'),
    alt.Y('co2:Q')
).properties(
    width=100,
    height=100
)

base.facet(
    facet=alt.Facet('country:N'),
    columns=6
)
```

B.
```
base = alt.Chart(df).mark_line().encode(
    alt.X('year:O'),
    alt.Y('co2:Q')
).properties(
    width=100,
    height=100
)

base.grid(
    row=alt.Row('country:N'),
    rows=6
)
```

C.
```
base = alt.Chart(df).mark_line().encode(
    alt.X('year:O'),
    alt.Y('co2:Q')
).properties(
    width=100,
    height=100
)

base.matrix(
    column=alt.Column('country:N'),
    row=alt.Row('co2:Q')
)
```

D.
```
base = alt.Chart(df).mark_line().encode(
    alt.X('year:O'),
    alt.Y('co2:Q')
).properties(
    width=100,
    height=100
)

base.repeated(
    facet=alt.Facet('country:N'),
    shape=(8,6)
)
```