FIT3179 Data Visualisation

Test 1 Sample Questions

Below are questions from tests of previous years. Test 1 of this year will not include questions consisting of text only, such as the first three questions below. However, these questions are still useful to understand the tested knowledge.

Some sample questions are multiple choice questions. Test 1 of this year will only include questions with a single correct answer.

The sample questions mainly focus on marks, channels, dataset types, attribute types and simple table idioms. However, you should also expect questions about data-ink ratio, chart junk, lying visualisation, storytelling, colour, visual hierarchy, typography, layout, as well as advanced idioms for table datasets, and idioms for networks and trees.

Answers to the sample questions can be found at the end of this document.

General concepts

1. **Q: Which of the following terms is equivalent to marks in data visualisation?**

A) Interaction primitives

B) Data semantic

C) Position and location

D) Colour channels

E) Geometry primitives

2. **Q: Which of the following terms is equivalent to visual channels in data visualisation?**

A) Colour channels

B) Data semantic

C) Position and location

D) Retinal variables

E) Geometry primitives

3. **Which of the following terms is equivalent to visual channels in data visualisation?**

A) Colour channels

B) Hue, saturation and brightness

C) Position and location

D) Visual variables

E) Point, line, area

Colour channels

4. **Select the main changes in the following colour scale**

A) Colour hue

B) Colour saturation

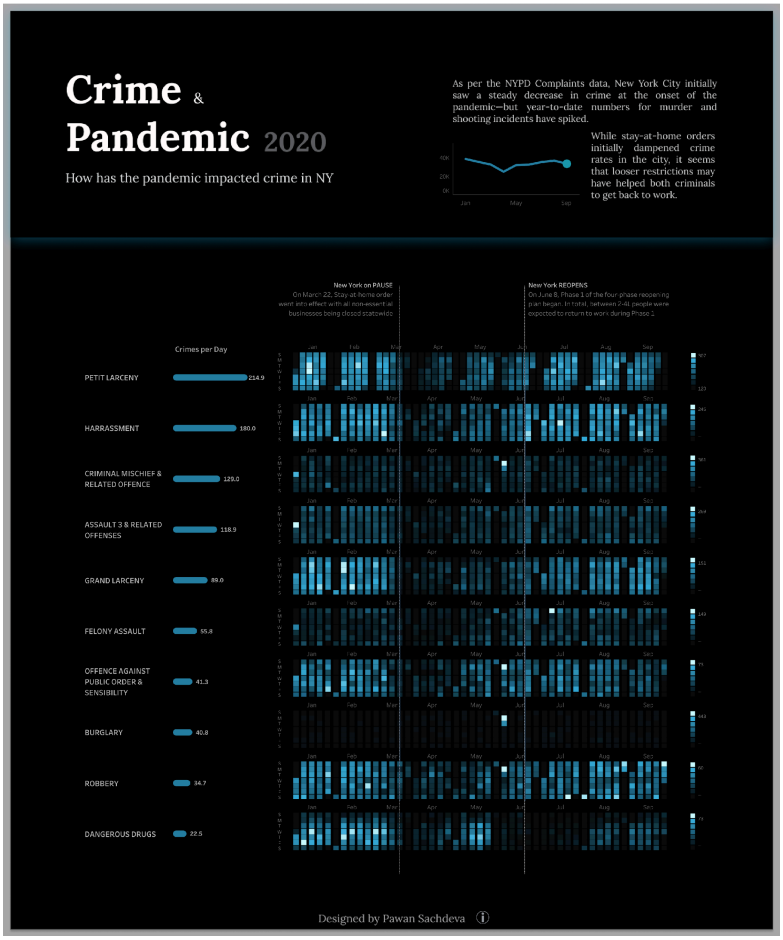
C) Colour luminance

5. **What is the colour channel used in the following visualisation? (hint: check the legend on the right side of the visualisation).**

A) Colour hue

B) Colour saturation

C) Colour luminance



Attribute type of a given attribute

6. **Let's consider a person's blood type (A, AB, B, O, etc.) as an attribute of a medical dataset. This attribute is best described as which one of the following?**

A) categorical attribute

B) ordinal attribute

C) quantitative attribute

D) sequential attribute

E) diverging attribute

7. **Let's think about temperature measurements in Celsius degrees as an attribute of a city dataset. This attribute is best described as which one of the following.**

A) categorical attribute

B) ordinal attribute

C) quantitative attribute

D) sequential attribute

E) qualitative attribute

8. **Let's consider the price range ("below 50 dollars", "50 to 100 dollars", "100 to 200 dollars", "above 200 dollars") as an attribute of a clothes dataset. This attribute is best described as which one of the following?**

A) categorical attribute

B) ordinal attribute

C) quantitative attribute

D) nominal attribute

E) diverging attribute

Given a dataset, select all attribute types.

The table below presents ten rows of a Victorian real estate dataset. The attributes include:

1) latitude

2) longitude

3) price of the property

4) type: house, apartment, townhouse, unit, etc.

5) bedroom: 1, 2, 3, 4, or 5+ ("5+" means 5 or more bedrooms)

6) travel_time_to_MC: Travel time to Melbourne Central by train (in minutes).

7) dist_to_supermarket: Distance to the nearest supermarket (in meters)

8) Suburb

9) Postcode

lat	lng	price	type	bedroom	travel_time_to_MC	dist_to_supermarket	Suburb	postcode
-37.9239	145.1698	1220000	House	3	36	1050	Mulgrave	3170
-37.7825	144.9898	560000	Apartment	2	25	1818	Fitzroy North	3068
-37.8342	144.9376	385000	House	2	4	632	Port Melbourne	3207
-37.8197	145.0973	500000	House	3	21	1462	Surrey Hills	3127
-37.8074	144.9608	560000	Apartment	2	4	384	Melbourne	3000
-37.8469	144.9787	692000	Apartment	2	12	266	Melbourne	3004
-37.7618	145.1944	465000	House	5+	31	2218	Warrandyte	3113
-37.7979	144.8077	650200	Townhouse	3	20	200	Sunshine West	3020
-37.7979	144.8241	736000	House	3	20	1637	Sunshine	3020
-37.7559	144.7514	820000	House	4	28	2160	Burnside	3023

9. **Please select all the categorical (nominal) attributes**

Note: Please select all correct answers.

A) lat

B) lng

C) Price

D) Type

E) Bedroom

F) Travel_time_to_MC

G) Dist_to_supermarket

H) Suburb

I) Postcode

10. **Select all the ordinal attribute(s):**

Note: Please select all correct answers.

A) lat

B) lng

C) Price

D) Type

E) Bedroom

F) Travel_time_to_MC

G) Dist_to_supermarket

H) Suburb

I) Postcode

11. **Select all the quantitative attributes.**

Note: Please select all correct answers.

A) Price

B) Type

C) Bedroom

D) Travel_time_to_MC

E) Dist_to_supermarket

F) Suburb

G) Postcode

Attribute types vs visual channels

12. **Which one of the following attributes is NOT suitable to be encoded with the LENGTH visual channel?**

A) The height attribute from a student dataset

B) The age attribute from a customer dataset

C) The nationality attribute from a student dataset

D) The land size attribute from a real estate dataset

13. **Which one of the following attributes is NOT suitable to be presented using the COLOUR SATURATION visual channel?**

A) The height attribute from a student dataset

B) The citizenship attribute from an employee dataset

C) The price attribute from a property dataset

D) The age attribute from a customer dataset

Channel ranking

14. **Let's think about visualising the property type attribute (house, unit, apartment, etc.) in a real estate dataset. Select the most effective visual channel from below to encode this attribute.**

A) Length

B) Shape

C) Area

D) Colour hue

E) Colour saturation

15. **Let's think about visualising a "birth rate" attribute in a country dataset. Select the most effective visual channel from below to encode this attribute.**

A) Length

B) Shape

C) Area

D) Colour luminance

E) Colour saturation

16. **In visualisation design, "the most important attributes should be encoded with the most effective visual channels in order to be most noticeable, and then decreasingly important attributes can be matched with less effective channels" (unit textbook by Tamara Munzner, 2014, p 101).**

Suppose that we use a scatter plot (or bubble chart) to visualise the covid-19 dataset below. Each point represents a country; the colour hue is used to represent "Continent". The importance ranks of the attributes that we would like to visualise are: "Active (cases)" > "Deaths" > "Confirmed (cases)".

Which of the following visualisation designs is the most effective?

Country	Confirmed	Deaths	Recovered	Active	Population	Continent
Afghanistan	39703	1473	33064	5166	38928341	Asia
Albania	15231	416	9406	5409	2877800	Europe
Algeria	52940	1795	37170	13975	43851043	Africa
Andorra	2696	55	1814	827	77265	Europe
Angola	6246	218	2716	3312	32866268	Africa
Antigua and Barbuda	111	3	97	11	97928	North America
Argentina	883882	23581	709464	150837	45195777	South America
Armenia	55736	1016	45771	8949	2963234	Europe
Australia	27263	898	24987	1378	25499881	Oceania
Austria	54423	852	42829	10742	9006400	Europe
Azerbaijan	41752	608	39235	1909	10139175	Europe
Bahamas	5023	106	2815	2102	393248	North America
Bahrain	75287	273	70808	4206	1701583	Asia
Bangladesh	377073	5500	291365	80208	164689383	Asia
Barbados	206	7	183	16	287371	North America
Belarus	82471	885	76543	5043	9449321	Europe
Belgium	156931	10175	20202	126554	11589616	Europe

A) Use x-axis to represent "Active cases", y-axis to represent "Confirmed cases", size (area) to represent "Deaths"

B) Use x-axis to represent "Confirmed cases", y-axis to represent "Active cases", size (area) to represent "Deaths"

C) Use x-axis to represent "Confirmed cases", y-axis to represent "Deaths", size (area) to represent "Active cases"

D) Use x-axis to represent "Active cases", y-axis to represent "Deaths", size (area) to represent "Confirmed cases"

Idiom vs attribute types

The table below presents ten rows of a Victorian real estate dataset. The attributes include:

1) latitude
2) longitude
3) price of the property
4) type: house, apartment, townhouse, unit, etc.
5) bedroom: 1, 2, 3, 4, or 5+ ("5+" means 5 or more bedrooms)
6) travel_time_to_MC: Travel time to Melbourne Central by train (in minutes).
7) dist_to_supermarket: Distance to the nearest supermarket (in meters)
8) Suburb
9) Postcode

lat	lng	price	type	bedroom	travel_time_to_MC	dist_to_supermarket	Suburb	postcode
-37.9239	145.1698	1220000	House	3	36	1050	Mulgrave	3170
-37.7825	144.9898	560000	Apartment	2	25	1818	Fitzroy North	3068
-37.8342	144.9376	385000	House	2	4	632	Port Melbourne	3207
-37.8197	145.0973	500000	House	3	21	1462	Surrey Hills	3127
-37.8074	144.9608	560000	Apartment	2	4	384	Melbourne	3000
-37.8469	144.9787	692000	Apartment	2	12	266	Melbourne	3004
-37.7618	145.1944	465000	House	5+	31	2218	Warrandyte	3113
-37.7979	144.8077	650200	Townhouse	3	20	200	Sunshine West	3020
-37.7979	144.8241	736000	House	3	20	1637	Sunshine	3020
-37.7559	144.7514	820000	House	4	28	2160	Burnside	3023

17. Select a pair of attributes that is suitable to be visualised with a scatter plot.

A) Price and Type

B) Price and Bedroom

C) Price and Dist_to_supermarket

D) Price and Suburb

E) Price and Postcode

18. Select all the attributes that are suitable to be visualised in a scatterplot matrix.

Note: Please select all correct answers.

A) price

B) type

C) travel_time_to_MC

D) dist_to_supermarket

E) Suburb

F) Postcode

19. **Assume we visualise all the properties as dots on a map based on the latitude and longitude attributes. Select all the magnitude channels that are suitable to encode the price attribute.**

Note: Please select all correct answers.

A) Area

B) Colour saturation

C) Colour hue

D) Shape

E) Colour luminance

F) motion

Visual channels (easy)

20. **Consider the following visualisation. What is the visual channel used to represent the Accessibility Remoteness Index Australia (ARIA+) (2006)? Hint: check the legend on the left-bottom corner.**

A) Area

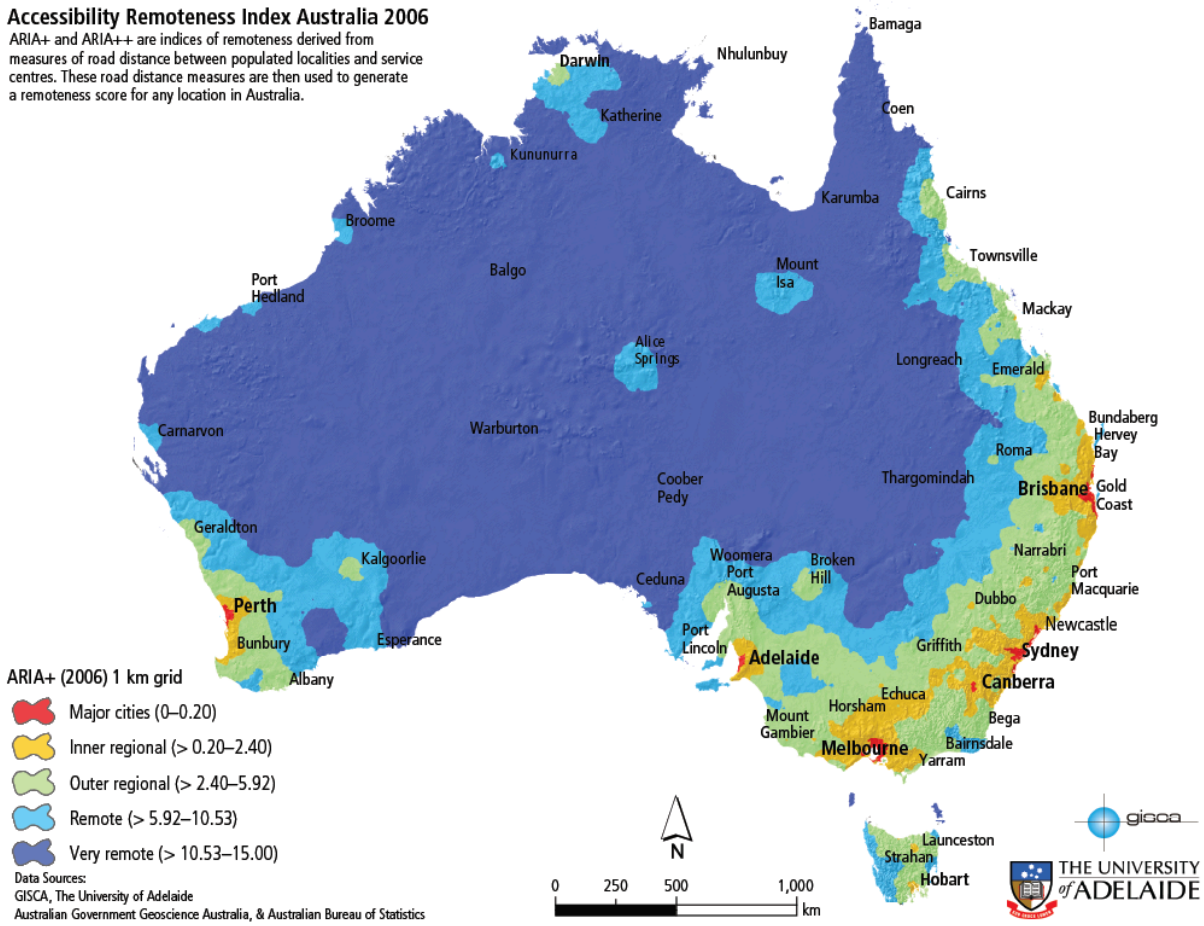
B) Size

C) Position

D) Colour hue

E) Colour saturation

F) Colour luminance



Visual Channels (difficult)

21. The horizon graph idiom is a variation of the line chart idiom, which can visualise time series data with much less space than line charts. The following graphs (Steps 1-4) present how a horizon graph is created.

Consider the final visualisation (Step 4) of this particular horizon graph.

Select all the channels used in this visualisation.

Note: Please select all correct answers.

A) Length

B) Position

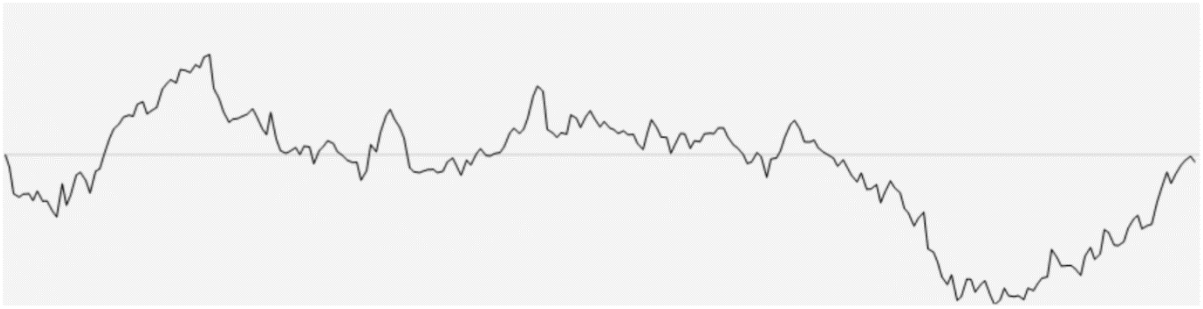
C) Area

D) Colour hue

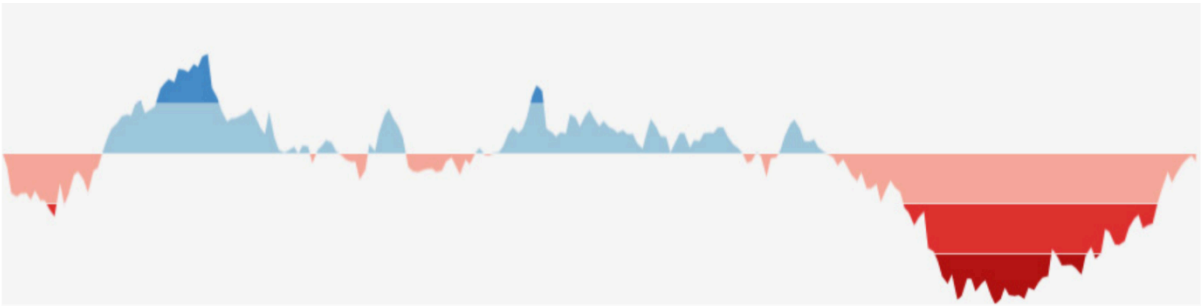
E) Colour saturation and/or luminance

F) Shape

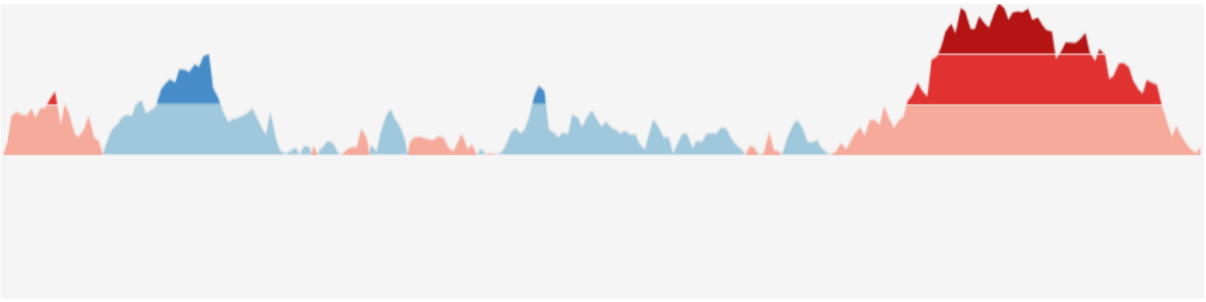
G) Texture



Step 1 - Create a line chart.



Step 2. Use variations in colour to make patterns and exceptions more visible.



Step 3. Display increases and decreases in value in a shared vertical space.

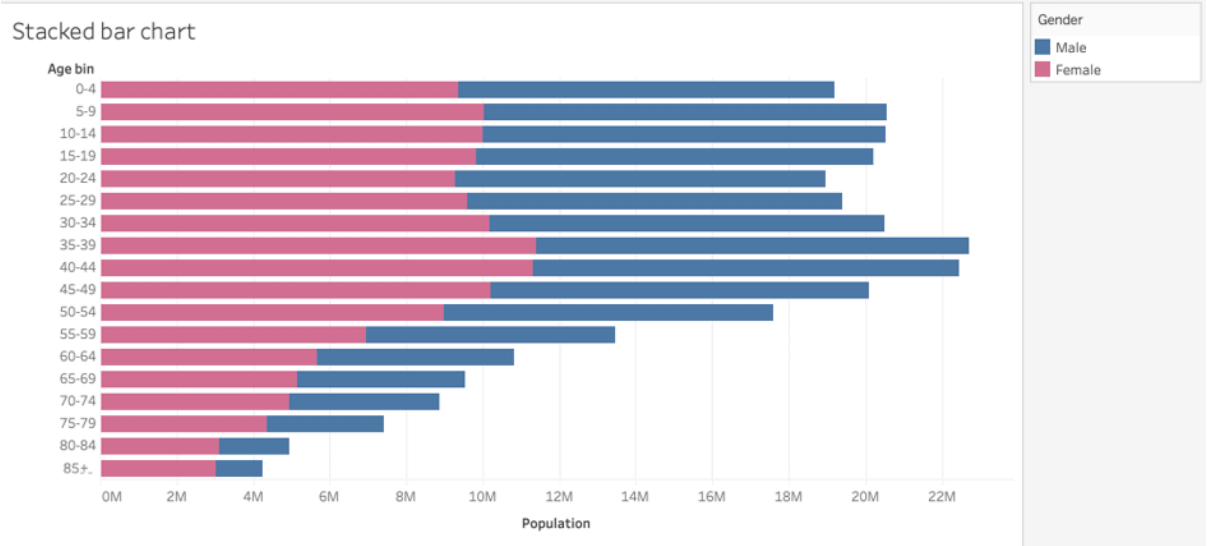
Step 4. Collapse the colour bands to display the values in less vertical space.

[Source: http://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf]

Compare bar chart and population pyramid (difficult)

22. Which of the following statements are correct? Select all correct answers.

Note: Please select all correct answers.



A) This graph shows two attributes.

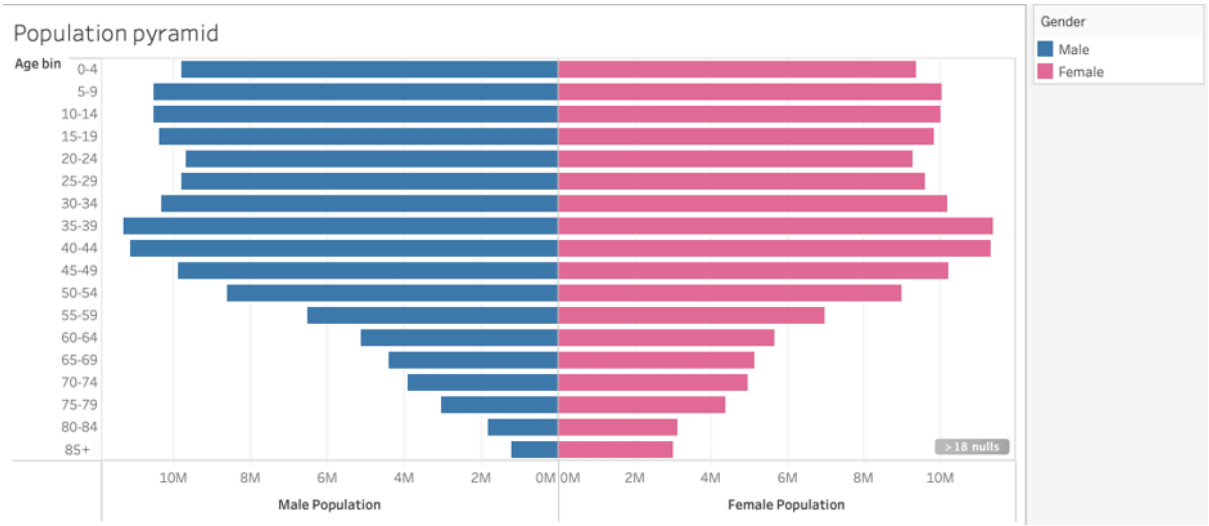
B) Gender is a nominal attribute, which is visualised with the colour hue channel.

C) The age bin (for example, “0–4”) is an ordered attribute that is shown with the position channel.

D) The marks used are lines and there are 20 million males in the age bin “45–49”.

23. Which of the following statements are correct? Select all correct answers.

Note: Please select all correct answers.



.

A) This graph shows two attributes.

B) Gender is a nominal attribute, which is visualised with both the colour hue channel and the “spatial region” channel.

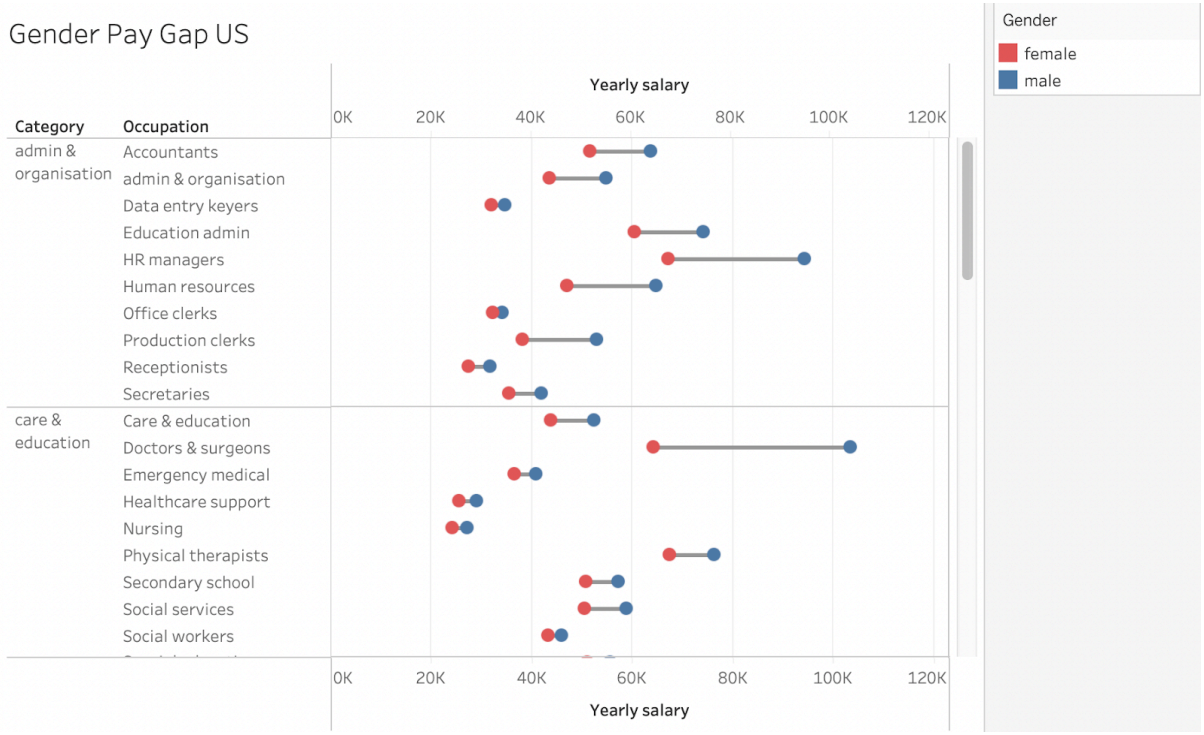
C) The age bin (for example, “0–4”) is an ordered attribute that is shown with the position channel.

D) The marks used are lines and there are about 20 million people in the age bin “45–49”.

Compare bar chart and connected dot plot (difficult)

24. This visualisation shows a Gender Pay Gap dataset for the USA. Which of the following statements are correct? Select all correct answers.

Note: Please select all correct answers.



A) This chart shows three categorical attributes, and one of these categorical attributes is shown with the colour hue channel.

B) The marks used for the quantitative attribute(s) are points and lines.

C) The salary of male accountants is about $65K higher than the salary of female accountants.

D) The horizontal line marks between the red and blue dots should be removed because they are distracting.

25. Which of the following is not a data attribute type?

A) Qualitative.

B) Ordinal.

C) Ratio.

D) Diverging.

E) Categorical.

26. Let's think about the size of T-shirts (XS, S, M, L, XL, etc.) as a variable. This variable is best described as which one of the following?

A) categorical attribute

B) ordinal attribute

C) quantitative attribute

D) numeric attribute

E) nominal attribute

27. Which of the following channels is the most effective to encode qualitative data attributes?

A) Length

B) Shape

C) Area

D) Colour hue

E) Colour saturation

28. The bubble chart below presents an overview of different countries in 2019. Each bubble represents a country, and the visual encodings used in the graph include:

*** x-axis position: income**

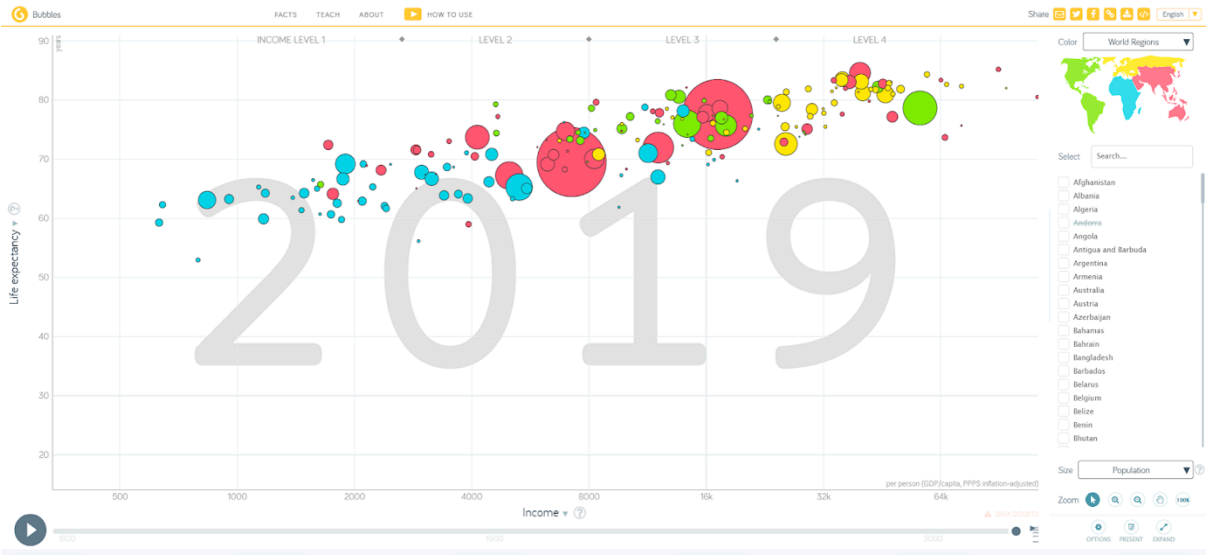
*** y-axis position: life expectancy**

*** area size: population**

*** colour hue: continent**

*** animation: year**

Which of the following list all the attributes that are represented with magnitude channels in this visualisation?



[Source: https://www.gapminder.org/tools/#$chart-type=bubbles]

A) income, life expectancy & population

B) income, life expectancy & continent

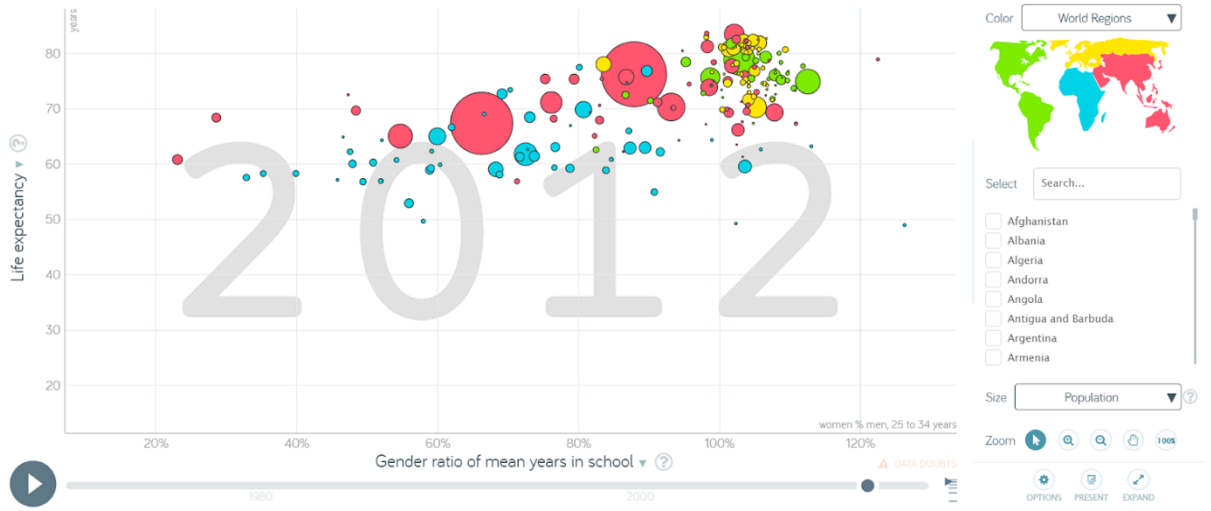
C) income & continent

D) population, continent & year

E) income, life expectancy

29. Consider the following bubble chart from Gapminder [link] which compares different countries in a few aspects (population, life expectancy, etc.). Select all the channels used in this visualisation.

(+1 mark per correct channel identified, -1 mark per incorrect channel selected).



A) Length

B) Position

C) 1D Size

D) Colour hue

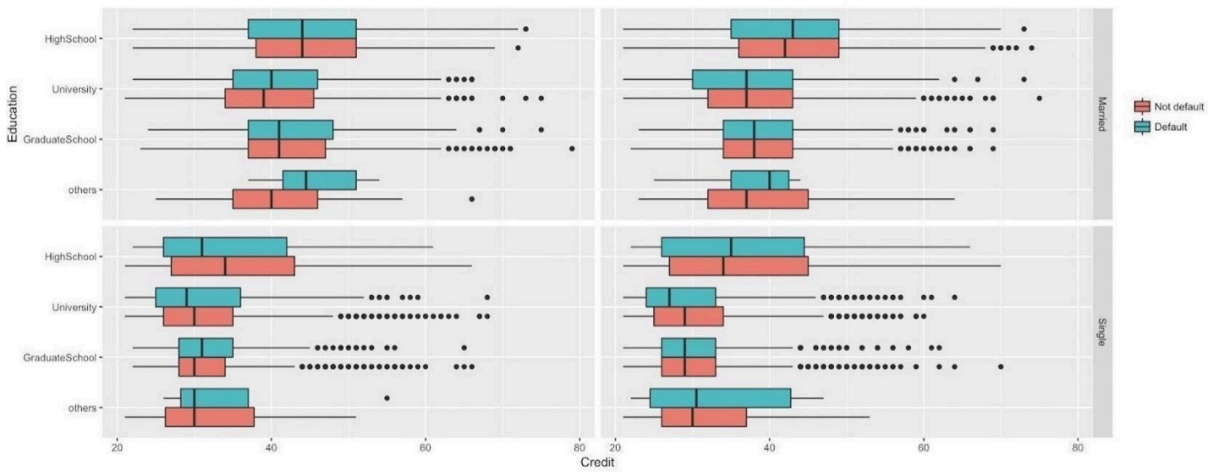
E) Colour Saturation

F) Colour luminance

G) Area

H) Shape

30. The following boxplot explores factors that may be associated with defaulting on loan repayment. List all the channels used in this visualisation.



A) Length

B) Position

C) Area

D) 2D Size

E) Colour hue

F) Colour Saturation

G) Colour luminance

H) Shape

31. Which of the following statements is not true regarding line charts?

A) The line chart idiom is suitable to show how values develop over time.

B) You can use colours, line width and line dashes in a line chart to make your most important values stick out.

C) Using annotations might make your line chart more interesting to read.

D) A line chart is more effective to show how values differ in different categories compared to a (stacked) bar chart.

E) It is a good idea to show each line in a line chart with a distinctive colour.

32. What is an appropriate idiom for a table dataset with two quantitative value attributes?

A) Scatter plot

B) Bar chart

C) Pie Chart

D) Area Chart

E) Line chart

Correct Answers

1. E
2. D
3. D
4. B
5. C
6. A
7. C
8. B
9. D, H, I
10. E
11. A, D, E
12. C
13. B
14. D
15. A
16. D
17. C
18. A, C, D

19. A, B, E
20. D
21. B, D, E
22. B, C
23. B, C, D
24. A, B
25. D
26. B
27. D
28. A
29. B, D, G
30. A, B, E
31. D
32. A