

ISS608 Visual Analytics – DataViz Makeover 5

Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

Data Visualisation Link (Tableau Online):

https://public.tableau.com/profile/denise.chua#!/vizhome/Dataviz_Makeover05/Dashboard1?publish=yes

a. Critiques of Data Visualisation

Clarity

S/N	Comments
1	<p>The purpose of this visualisation is not conveyed well in the title of the chart. The title misleads readers into thinking that the data shown corresponds to actual reproducibility of research results, and that the bottom line is at least 70% of research is reportedly reproducible.</p> <p>In fact, this visualisation aims to show the percentage of respondents that believe at least 70% of the research in their area of interest is reproducible.</p>
2	The x-axis label, 'At least 70%', is not indicative of what the data values on the bar chart mean. Worse still, the mention of 70% may even mislead readers to think the values are related to the actual reproducibility of published research, as incorrectly insinuated in the title.
3	<p>The sorting of horizontal bars by decreasing proportion of respondents is useful to direct readers' attention on the top-ranked area of interests.</p> <p>However, adding additional information on the absolute number of respondents can provide more insights on the sample size of each sub-group of respondents, which may be important for readers to gauge the statistical significance level of the findings.</p>
4	The static visualisation solely provides the statistics for respondents that selected 70%,80%,90% and 100% (proportion of research believed to be reproducible), and therefore discards statistics on the other responses (0%-60%). It would be more informative if the distribution of responses across the range of available options was provided.

Aesthetic

S/N	Comments
1	<p>The use of color-coding here is presumably used to visually partition out the different areas of interest, but is superfluous in this case since the y-axis has labels which clearly indicate which bars belong to which category.</p> <p>Moreover, the use of colors should help to enhance the conveyance of information, but in this case, too many contrasting colors due to the presence of too many categories makes it harder to read the data.</p>
2	The legend, which should not appear in the visualisation in the first place (as explained in the above comment), is used ineffectively because it is not sorted by descending order unlike the bar chart. This makes it harder for readers to map the colors to the corresponding category.
3	Units are not indicated on x-axis tick values, and may confuse readers into thinking that it is an average value rather than percentage of total.

b. Ways to improve current design

Clarity

Critiques (from part a)	Suggested Improvement
The purpose of this visualisation is not conveyed well in the title of the chart. The title misleads	Dashboard title can be modified to "CONFIDENCE IN REPRODUCIBILITY OF

ISSS608 Visual Analytics – DataViz Makeover 5

Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

<p>readers into thinking that the data shown corresponds to actual reproducibility of research results, and that the bottom line is at least 70% of research is reportedly reproducible.</p> <p>In fact, this visualisation aims to show the percentage of respondents that believe at least 70% of the research in their area of interest is reproducible.</p>	<p>PUBLISHED RESEARCH” to reflect that the visualization is the confidence level of researchers/ what the researchers believe is the reproducibility of the research in their area of interest.</p> <p>The chart title should incorporate the statement: “How Many Researchers Believe That At Least 70% of Research In Their Field Is Reproducible?”</p>
<p>The x-axis label, ‘At least 70%’, is not indicative of what the data values on the bar chart mean. Worse still, the mention of 70% may even mislead readers to think the values are related to the actual reproducibility of published research, as incorrectly insinuated in the title.</p>	<p>The x-axis labels will be modified to “%respondents that believe %X of research in their field is reproducible” to reflect the meaning the data values on the bar chart. Note: X% refers to a dynamic value that the user can indicate, with the use of parameter feature in tableau</p>
<p>The sorting of horizontal bars by decreasing proportion of respondents is useful to direct readers’ attention on the top-ranked area of interests.</p> <p>However, adding additional information on the absolute number of respondents can provide more insights on the sample size of each sub-group of respondents, which may be important for readers to gauge the statistical significance level of the findings.</p>	<p>To retain the sorting of horizontal bars by decreasing proportion of respondents. A histogram/ count plot of number of responses across the survey response options will be created, and added into the tooltip of the lollipop chart. This allows users to view the distribution and absolute count of responses, based on detail-on-demand.</p>
<p>The static visualisation solely provides the statistics for respondents that selected 70%,80%,90% and 100% (proportion of research believed to be reproducible), and therefore discards statistics on the other responses (0%-60%). It would be more informative if the distribution of responses across the range of available options was provided.</p>	<p>To provide users the ability to select other survey response options (0% - 60%), and also the ability to select “at least”, “at most”, and “exactly”, through the use of parameters. Based on the users’ select input, the lollipop chart will vary as the data is computed differently, potentially revealing more insightful information from the dynamic visualisation</p>

Aesthetic

Critiques (from part a)	Suggested Improvement
<p>The use of color-coding here is presumably used to visually partition out the different areas of interest, but is superfluous in this case since the y-axis has labels which clearly indicate which bars belong to which category.</p> <p>Moreover, the use of colors should help to enhance the conveyance of information, but in this case, too many contrasting colors due to the</p>	<p>Instead of using areas of interest to color-code the categories, the % of respondents data values will determine the lightness/darkness of a color through the use of gradient palette. This puts emphasis on the research areas with high confidence levels of results reproducibility.</p>

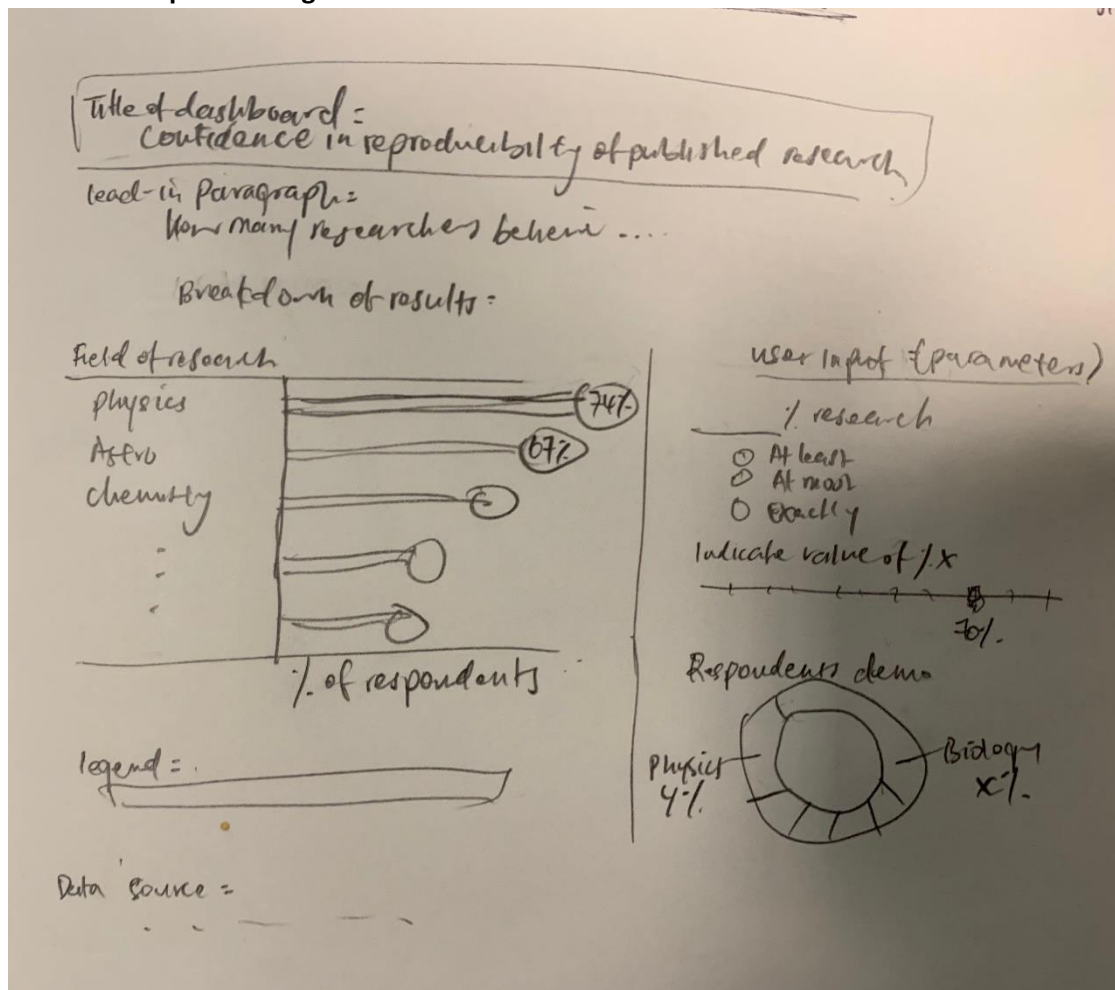
ISSS608 Visual Analytics – DataViz Makeover 5Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

presence of too many categories makes it harder to read the data.	The legend, in this case, would be placed in the visualization for readers to validate the visual encoding.
The legend, which should not appear in the visualisation in the first place (as explained in the above comment), is used ineffectively because it is not sorted by descending order unlike the bar chart. This makes it harder for readers to map the colors to the corresponding category.	
Units are not indicated on x-axis tick values, and may confuse readers into thinking that it is an average value rather than percentage of total.	Units (%) will be indicated in all the tick values on the x-axis

ISS608 Visual Analytics – DataViz Makeover 5

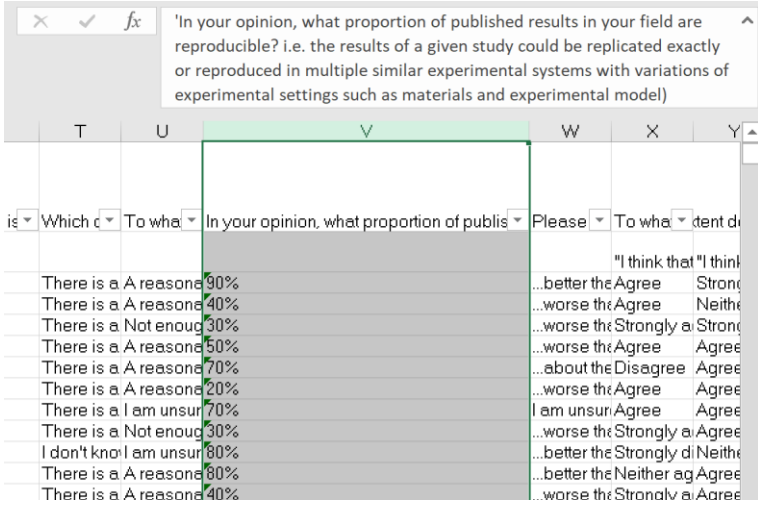
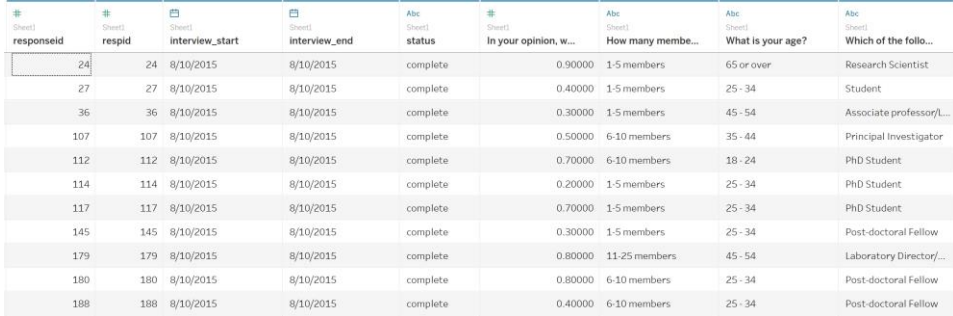
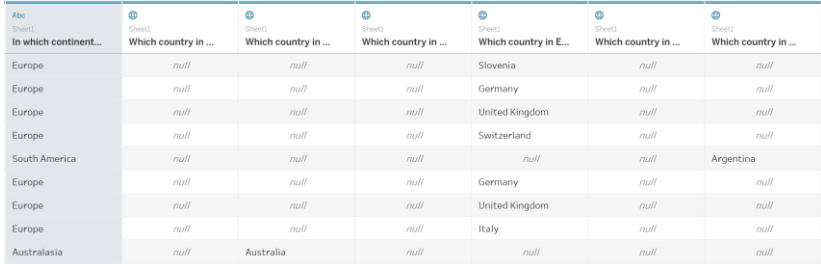
Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

Sketch of Proposed Design



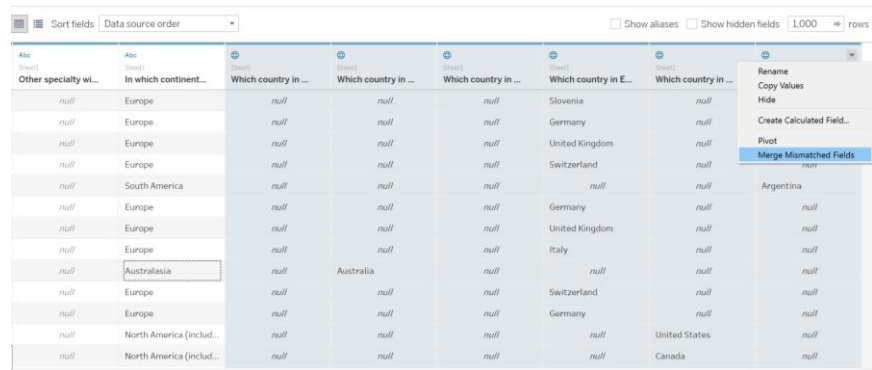
- To retain the original intent of the visualisation, lollipop chart (which is a variant of horizontal bar chart) sorted by largest to smallest proportion of respondents was used to put emphasis on the top-ranked area of interest; While the lollipop circle is used as a visual element to highlight data value, which is what the readers are seeking after.
- The use of displaying histogram inside the tooltip box is to provide readers details-on-demand if they are interested to know the distribution of responses within each area of interest. The histogram also provides readers additional information on the actual count of responses.
- Parameter features will be available for users to indicate which survey response options (0% -100%) they would like to see the results of; And users with the ability to select “at least”, “at most”, and “exactly”. Based on the users’ select input, the lollipop chart will vary as the data is computed differently, potentially revealing more insightful information from the dynamic visualisation

c. Step-by-step description on data visualization preparation

Tab	Step	Action
Data Source	Import to tableau	<p>Inspecting the raw excel data file</p> <ul style="list-style-type: none"> Column V correspond to the survey responses that are pertinent to the original visualisation intent to show “What proportion of published results in your field are reproducible?”. Columns F:U and W:CH correspond to survey responses for questions that are irrelevant and therefore were deleted in the excel sheet so as to obtain a question-focused dataset to work with in Tableau. Columns CI:DK correspond to characteristics of respondents/ demographic data (including area of interest), are potentially useful and therefore kept first for further analysis.  <p>Importing modified excel file (.xlsx) into tableau:</p> <ul style="list-style-type: none"> It was verified that the headers and data rows were correctly interpreted by tableau upon connecting the data source to the excel file. 
Additional data cleaning and shaping		<ul style="list-style-type: none"> Some demographics data, such as country of origin, are separated into a few columns in the original dataset, therefore contains many null values 

ISS608 Visual Analytics – DataViz Makeover 5
 Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

To combine the data into one single column, the relevant columns were selected, and ‘Merge Mismatch Fields’ was selected from the dropdown menu.



The resulting column:

In which continent...	Which country in ...
Europe	Slovenia
Europe	Germany
Europe	United Kingdom
Europe	Switzerland
South America	Argentina
Europe	Germany
Europe	United Kingdom
Europe	Italy
Australasia	Australia

The same steps were performed to merge columns pertinent to “What is your specialty...” and “Other specialty...” into a single column each.

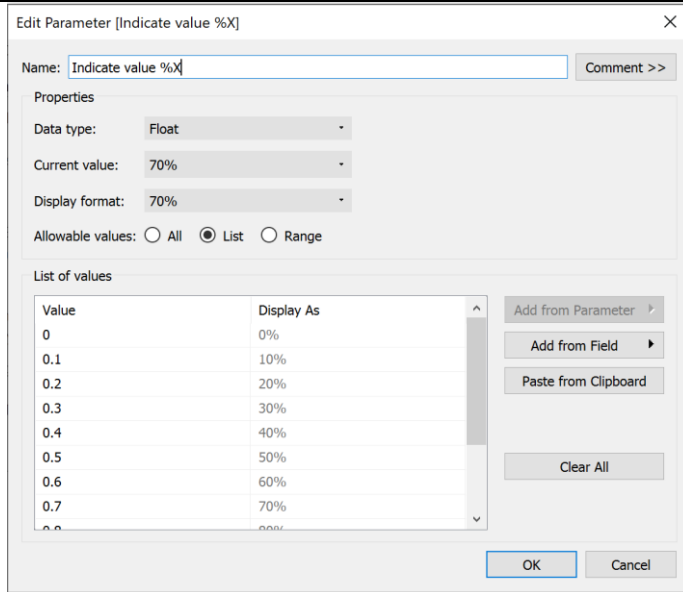
- Some columns were renamed to a more intuitive and concise caption. Specifically,
 1. “Which of the following best describes your area of interest?” was renamed to “Field of research”
 2. “In your opinion, what proportion of published results in your field are reproducible? i.e. the results of a given study could be replicated exactly or reproduced in multiple similar experimental systems with variations of experimental settings such as materials and experimental model)” was shortened to “In your opinion, what proportion of published results in your field are reproducible?”

Sheets 1 and 2

Creating new calculated fields and parameter for user selection and dynamic display

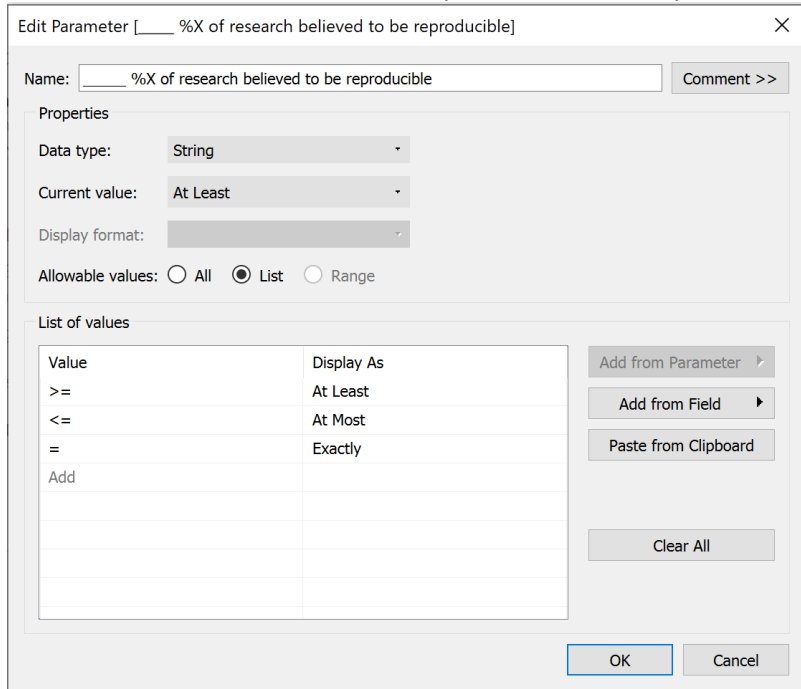
- In order to automatically compute proportion of respondents dynamically based on users’ selection, the following parameters were created
 1. [Indicate value of %X] parameter was first created. This parameter allows users to view the proportion of respondents that indicated %X of research is reproducible in their opinion, where X is the range of options (0% - 100%) in the survey.

ISS608 Visual Analytics – DataViz Makeover 5
 Denise Chua (deniseadele.2019@mitb.smu.edu.sg)




- The list of values was imported by selecting [In your opinion, what proportion of published results in your field are reproducible?] in the drop-down menu from “Add from Field” button.
- Data type was set to Float
- The display format was set to percentage, 0 decimal place
- The current value set to 70%

2. The second parameter [____ %X of research believed to be reproducible] was created, to allow users to view proportion of respondents that selected options which are <= %X, >= %X, or exactly %X, for which the value of X can be specified in the first parameter.


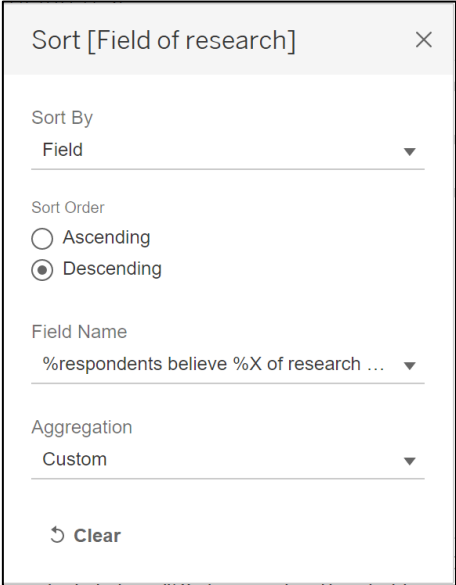


- The list of values are manually input, as shown in the screenshot above
- The data type is set as String

ISS608 Visual Analytics – DataViz Makeover 5
 Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

		<p style="text-align: center;">- Current value is set at “At Least”</p> <p style="text-align: center;">Both parameters are displayed by selecting “Show Parameter Control” in the right-click dropdown field.</p> <ul style="list-style-type: none"> A new calculated field [%respondents believe %X of research in their field is reproducible] was created. The formula includes several conditional statements, specifically values which users selected from the 2 parameters, to compute the proportion of respondents to be displayed in the lollipop chart. The exact formula used is given below: <pre> CASE [_____ %X of research believed to be reproducible] WHEN "<=" then SUM(IF [In your opinion, what proportion of published results in your field are reproducible] <= [Indicate value %X] then 1 else 0 END)/ SUM({Exclude [In your opinion, what proportion of published results in your field are reproducible (dimension)]: SUM([Number of Records])}) WHEN ">=" then SUM(IF [In your opinion, what proportion of published results in your field are reproducible] >= [Indicate value %X] then 1 else 0 END)/ SUM({Exclude [In your opinion, what proportion of published results in your field are reproducible (dimension)]: SUM([Number of Records])}) WHEN "=" then SUM(IF [In your opinion, what proportion of published results in your field are reproducible] = [Indicate value %X] then 1 else 0 END)/ SUM({Exclude [In your opinion, what proportion of published results in your field are reproducible (dimension)]: SUM([Number of Records])}) END </pre> <p>The use of {Exclude} LOD expression is to ensure that the computation is done on a [Field of Research] level, and not for individual survey response option.</p>  <p>The screenshot shows the Tableau calculated field editor for the field 'arch in their field is reproducible'. The formula is identical to the one provided in the text above. The editor includes a title bar, a close button (X), and a scrollable text area containing the code. At the bottom, it indicates 'The calculation is valid.', '4 Dependencies', and 'Apply' and 'OK' buttons.</p>
--	--	--

ISS608 Visual Analytics – DataViz Makeover 5
 Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

	<p>Creating lollipop chart</p>	<ul style="list-style-type: none"> [Field of research] was placed on Rows shelf, and [%respondents believe %X of research in their field is reproducible] was placed twice on Columns shelf. Both fields were automatically aggregated (as a result of the use of table calculation in the formula).  <ul style="list-style-type: none"> The second [AGG(%respondents believe %X of research in their field is reproducible)] mark was changed to 'dual axis', by selecting the option in the right-click dropdown menu. This automatically changes both marks to circle chart. The first [AGG(%respondents believe %X of research in their field is reproducible)] mark was changed to bar chart, the color changed to grey, and the size was reduced, so as to create skinny grey bars that resemble lollipop stick. The first [AGG(%respondents believe %X of research in their field is reproducible)] mark remained as circle chart, and the size was enlarged to resemble lollipop. [AGG(%respondents believe %X of research in their field is reproducible)] field was Ctrl+click and dragged onto color, so that higher data values are colored darker blue and lower data values are colored lighter blue, for visual encoding. [AGG(%respondents believe %X of research in their field is reproducible)] field was also dragged onto label, so that the data values are displayed on the circles. The second axis at the top of the chart was removed by deselecting "Show header" in the right-click dropdown menu. Finally the [Field of Research] was sorted in descending order of [AGG(%respondents believe %X of research in their field is reproducible)]. 
--	--------------------------------	--

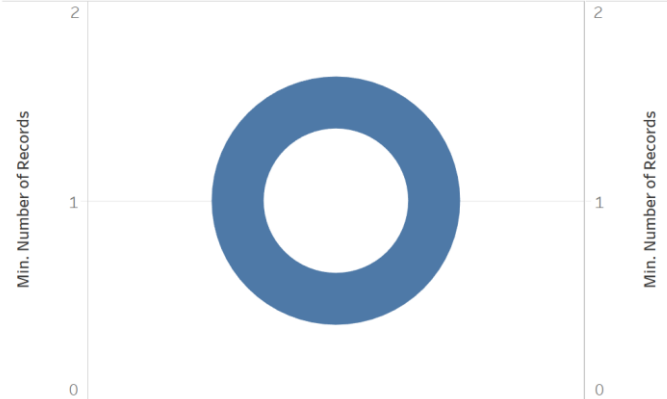
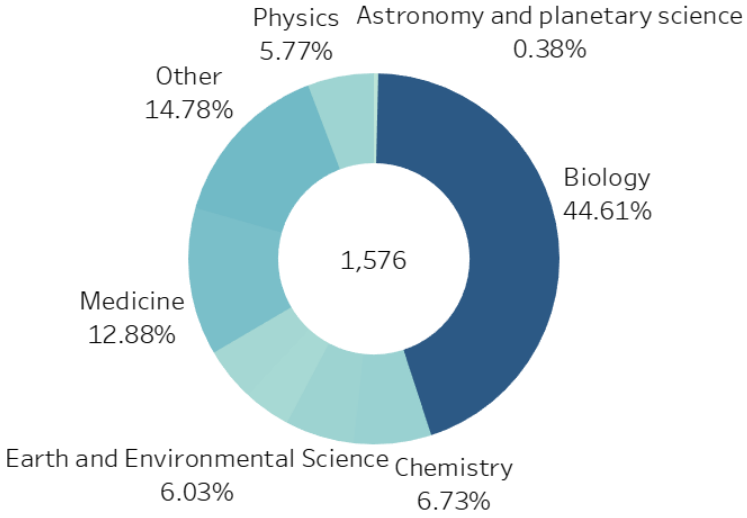
ISS608 Visual Analytics – DataViz Makeover 5

Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

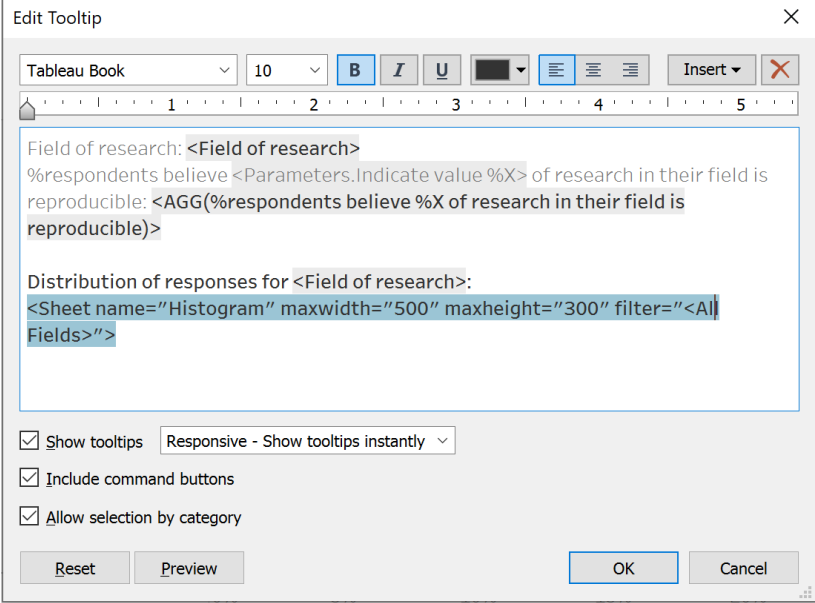
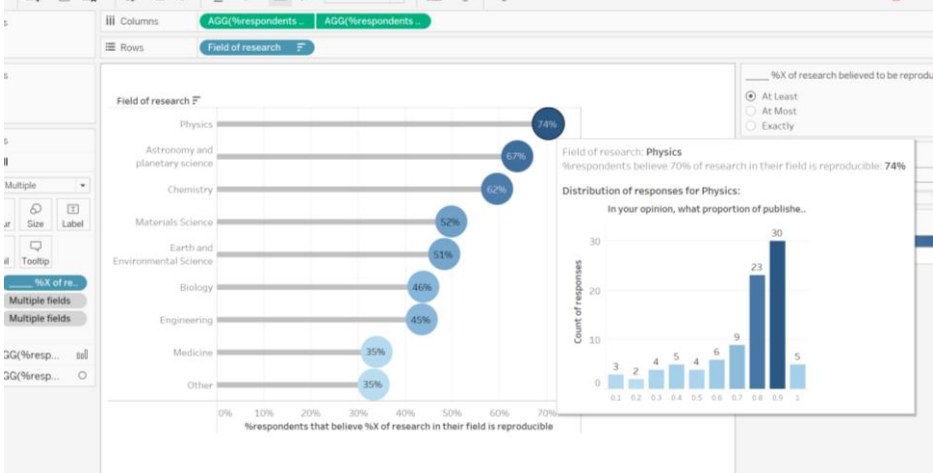
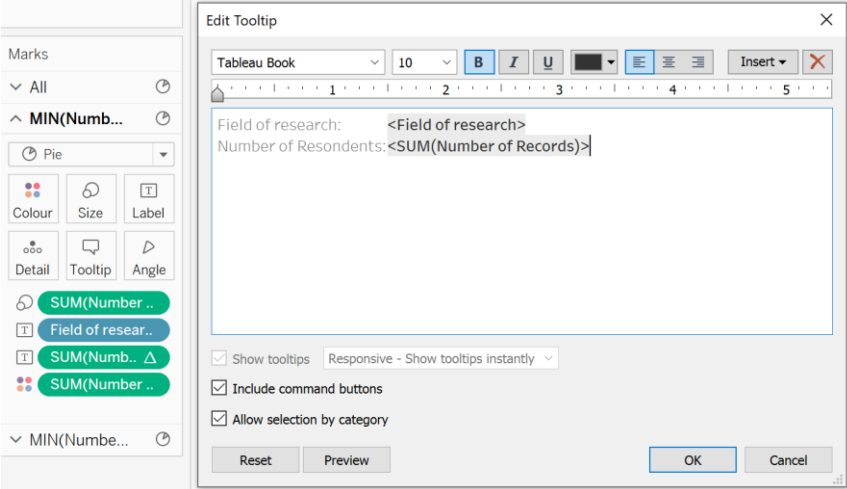
	<p>Setting up histogram (to be included in lollipop chart tool tip)</p>	<ul style="list-style-type: none"> [In your opinion, what proportion of published results in your field are reproducible] was placed both on Rows shelf and Columns shelf, and the field is automatically aggregated by SUM(). For the field on Rows shelf, the measure aggregation was changed to 'Count'. The field on Columns shelf was changed from measure to dimension, and the data type was changed from continuous to discrete COUNT([In your opinion, what proportion of published results in your field are reproducible]) field was Ctrl+click and dragged onto color, so that higher data values are colored darker blue and lower data values are colored lighter blue, for visual encoding. COUNT([In your opinion, what proportion of published results in your field are reproducible]) field was also dragged onto label, so that the data values are displayed on the bars. Y-axis title was changed to 'Count of responses' <p>Note: the tooltip(Field of research) field in the Filters pane will only appear after the histogram sheet is linked to the tooltip in the lollipop chart.</p>
	<p>Donut chart</p>	<ul style="list-style-type: none"> Under the Marks pane, pie was selected, [Field of research] was dragged to color and label

ISS608 Visual Analytics – DataViz Makeover 5

Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

		<ul style="list-style-type: none"> [Number of Records] was placed on the Rows shelf twice, and the aggregation of measure was changed from SUM to MIN for both fields The second [MIN(Number of Records)] mark was changed to 'dual axis', by selecting the option in the right-click dropdown menu. The color was changed to white, and the size was reduced. This creates the shape of a donut.  <p>[Number of Records], automatically aggregated to SUM, was placed onto Text</p> <ul style="list-style-type: none"> For the first [MIN(Number of Records)] mark and [Field of research] was placed on text. [Number of Records], automatically aggregated to SUM, was placed on color and size. [Number of Records] was manually changed to aggregate by percentage of total using quick table calculation and placed on Text. 
	<p>Tooltips and additional formatting</p>	<ul style="list-style-type: none"> The tooltip for All mark was edited the display the following Detail-on-Demand, which includes the link to the histogram chart.

ISSS608 Visual Analytics – DataViz Makeover 5
 Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

		 <p>How to tooltip looks like upon hovering over chart:</p>  <ul style="list-style-type: none"> • The tooltip for the first MIN(Number of Records) mark was edited as shown in the screenshot below 
<p>Dashboard</p>	<p>Create Dashboard, Edit titles</p>	<ul style="list-style-type: none"> • Create a new dashboard sheet. Sheet 1 and Sheet 2 were renamed 'Lollipop' and 'Donut' respectively and dragged into the dashboard sheet.

ISSS608 Visual Analytics – DataViz Makeover 5

Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

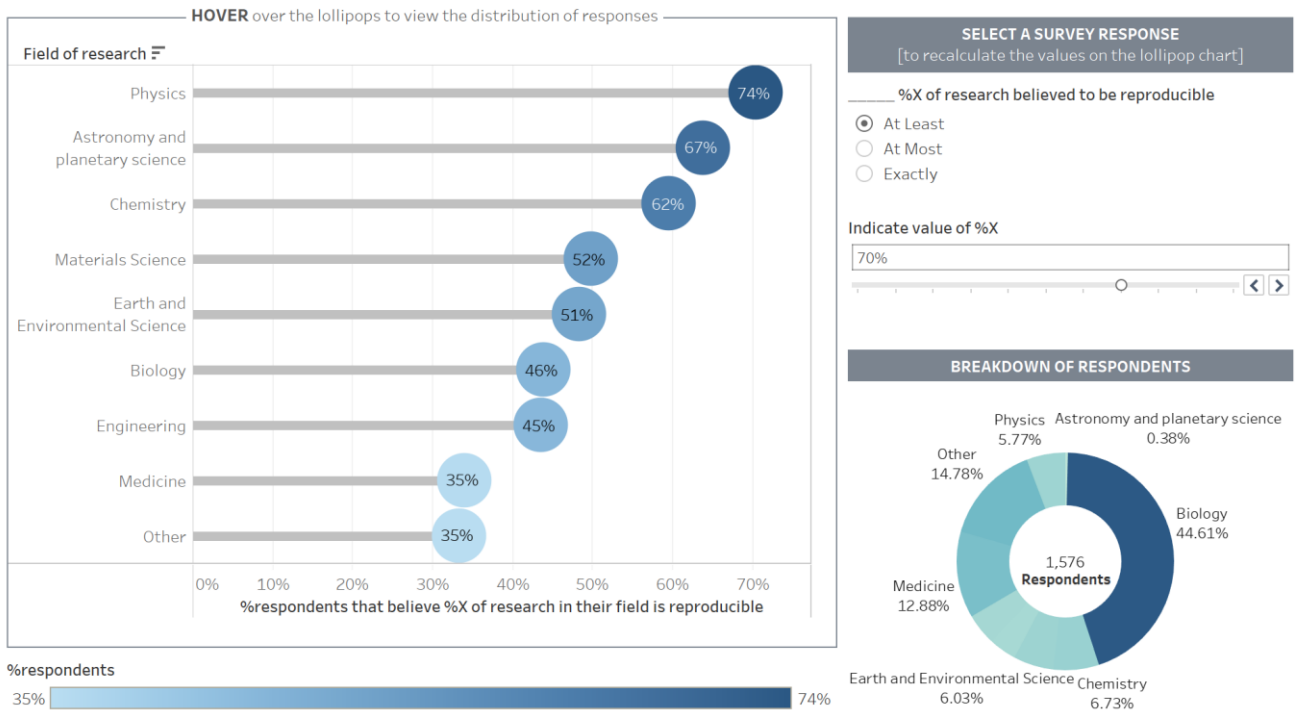
	<p>and Add additional elements</p>	<ul style="list-style-type: none"> • The title of the dashboard was changed to ‘CONFIDENCE IN REPRODUCIBILITY OF PUBLISHED RESEARCH’, with grey background and white text. • A textbox element was placed below the dashboard title with the following text to provide context to the purpose of the overall data visualization: <p style="text-align: center;">How Many Researchers Believe That</p> <p style="text-align: center;"><Parameters.____ %X of research believed to be reproducible></p> <p style="text-align: center;"><Parameters.Indicate value %X> of Research In Their Field Is Reproducible?</p> <p style="text-align: center;">1,576 researchers were surveyed, and the responses varied significantly across different fields of research.</p> <p style="text-align: center;">Here is the breakdown:</p> • A textbox was placed on top of the lollipop chart to provide instructions to users to “HOVER over the lollipops to view the distribution of responses” • A textbox was added on top of the parameters to direct users to “SELECT A SURVEY RESPONSE”.
--	------------------------------------	--

Dashboard Screenshot

CONFIDENCE IN REPRODUCIBILITY OF PUBLISHED RESEARCH

How Many Researchers Believe That **At Least 70%** of Research In Their Field Is Reproducible?

1,576 researchers were surveyed, and the responses varied significantly across different fields of research.
 Here is the breakdown:



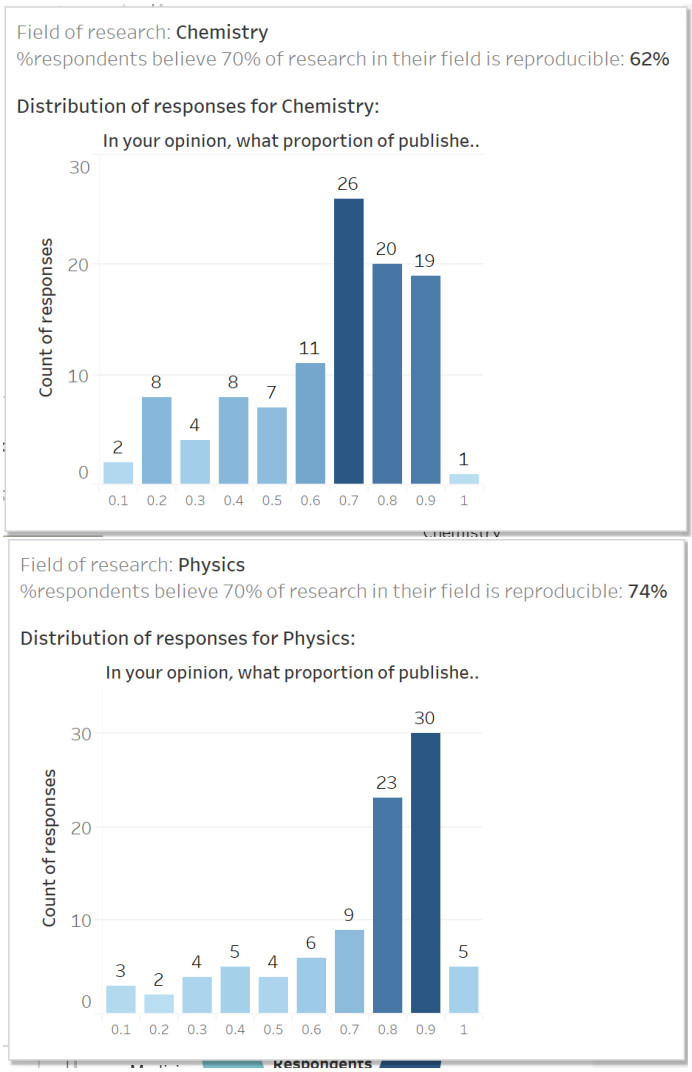
Source

1. Original article: <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>.
2. The raw survey data: https://www.nature.com/news/polopoly_fs/7_367421/file/Reproducibility%20Survey%20Raw%20Data.xlsx.
3. The survey questions: https://www.nature.com/news/polopoly_fs/7_367411/file/Reproducibility%20Questionnaire.doc

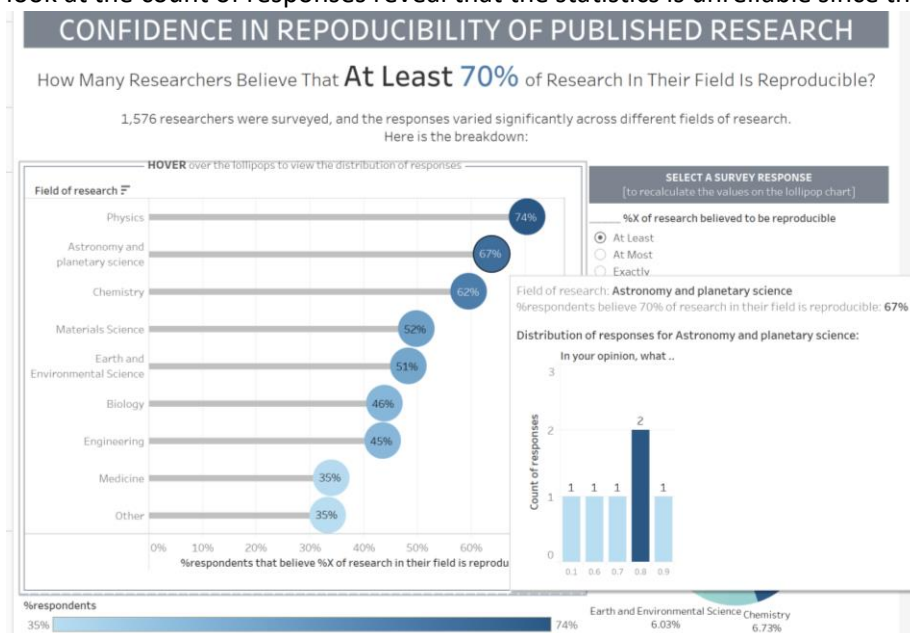
d. Information Revealed by Proposed Data Visualisation

1. The confidence level that respondents show in the reproducibility of published work in their field varies significantly across each field of research.
2. Physics and chemistry have significantly higher count of responses in the 70%/80%/90% survey options (right-skewed distribution); This means that physicists and chemists are generally more confident of their published research.

ISS608 Visual Analytics – DataViz Makeover 5
 Denise Chua (deniseadele.2019@mitb.smu.edu.sg)



3. While Astronomy and planetary science appears second-ranked in terms of proportion of respondents that believe at least 70% of the published research in the field is reproducible, a closer look at the count of responses reveal that the statistics is unreliable since the sample size is too small



ISS608 Visual Analytics – DataViz Makeover 5
 Denise Chua (deniseadele.2019@mitb.smu.edu.sg)

- On the reverse scale, 'Other', 'Earth and environmental science' and 'engineering' appear top 3 for proportion of researchers believe that at most 30% of published research in their field is reproducible, indicating low levels of confidence.

CONFIDENCE IN REPRODUCIBILITY OF PUBLISHED RESEARCH

How Many Researchers Believe That **At Most 30%** of Research In Their Field Is Reproducible?

1,576 researchers were surveyed, and the responses varied significantly across different fields of research. Here is the breakdown:

