Design & Implementation of 2 Visualisation Systems for Hiking Trails Data

The Residuals

Table of Contents

[I. Video Demonstration 3](#_Toc78027097)

[1. Data Set Description 3](#_Toc78027098)

[2. Visualisation Objectives 3](#_Toc78027099)

[3. System Implementation 3](#_Toc78027100)

[4. Design Comparison 3](#_Toc78027101)

[5. User Evaluation 4](#_Toc78027102)

[6. Future Work 4](#_Toc78027103)

[A. Appendix – Evaluation Data 5](#_Toc78027104)

[Bibliography 8](#_Toc78027105)

# Video Demonstration

Video demonstration of each visualisation system can be found at the following link.

**The Zebra system:** <https://youtu.be/bTgVpz0szy8>

**The Giraffe system:** <https://youtu.be/W0SNgEarHJw>

# Data Set Description

The Washington Hiking data set contains data on hiking trails in Washington State. The data set comes from the Tidy Tuesday community repository [1] and can be found at <https://github.com/rfordatascience/tidytuesday/tree/master/data/2020/2020-11-24>. It was obtained by scraping the Washington Trails Association website in November 2020 [2], after scraping the data set is offline. The data set type is a 2-dimensional table, with static availability. The data item is a hiking trail, which is specified by name. The data attributes are location, length, gain, highpoint, user rating, and features. Hike trail description is of text data type [3].

|  |  |  |
| --- | --- | --- |
| Attribute | Attribute Type | Ordering Direction |
| location | categorical | - |
| length | quantitative | sequential |
| gain | quantitative | diverging |
| highpoint | quantitative | diverging |
| features | categorical | - |

Table : Classification of Washington Hiking Data Attributes

# Visualisation Objectives

The main audience of the visualisation system is a beginner hiker visiting a national park in Washington State. The beginner hiker does not have a particular hiking trail in mind, and they would like to identify which one to take. The hiker would like to browse the available hiking trails at glance or look up which hiking trails have the attributes they are interested in.

The main outcome of the data visualisation is for hiker to know the name of the trail they want to hike. The target attributes are length and highpoint. For example, the hiker might like to know which trails are the shortest or provide the highest peak. The combination of these attributes is relevant, for example hight highpoint relative to short length would be challenging.

Hikers might also be interested in additional attributes, such as user rating or features of the trail. Rating should be conveyed by colour or size. In this way, hikers could also explore correlation between these attributes. For example, they could discover that longer trails with high peaks are lower rated. Lastly, the system could summarise facts about the region or trail features.

# System Implementation

The visualisation systems are named the Zebra System and the Giraffe System. Both can be viewed in the program code, which can be found in the accompanying folders **zebra.zip** and **giraffe.zip**. For demonstration, see Video Demonstration.

# Design Comparison

[6 decisions, 100 words each]

# User Evaluation

The evaluation was performed via an online survey. There were 5 participants in total. There were 3 females and 2 males in their 20s. For the Zebra visualisation, all participants were able to select a region using the dropdown menu. 4 participants were able to correctly identify the longest trail and the highest highpoint trail. All participants said there was no relationship between length and highpoint. These are subjective, however interesting given that linear relationships are better perceived using scatterplots. All participants would be able to select a hiking trail that interests them based on Zebra. Figure 1 shows overall rating participants gave to Zebra, the median rating was 4. Participants liked the straightforward layout, clearly displayed units of measurement, not too many data points plotted, and that it is simple to gain information on length vs. highpoint. They disliked that the points are too small and hard to select, little information beyond what is shown, that it requires concentration especially to new users.

For the Giraffe visualisation, 3 participants were able to correctly identify the longest and lowest highpoint trail. 1 participant noted that there were multiple options, since bins hide actual length and highpoint. 4 participants were able to identify the feature of the longest trail. Figure 2 shows Giraffe led participants to discover more relationships in the data. All participants would be able to choose a hiking trail of interest based on Giraffe. Figure 3 shows the overall rating participants gave to Giraffe, the median rating was 3. Participants liked that colour scheme, the simplicity of the graph, with more information given. They disliked that the axes are too small, graph was harder to make sense of, not visually appealing.

Participants thought Zebra is better at depicting the relationship in the data (60%), they would prefer seeing the features listed (80%). Giraffe was preferred by 3 participants, Zebra by 1 and neither by 1 (see Figure 4). Finally, 1 participant recommended not to use Giraffe (too confusing), 1 participant preferred Giraffe since Zebra might not work on mobile, and another recommended to make dimensions consistent, with bigger point and text size.

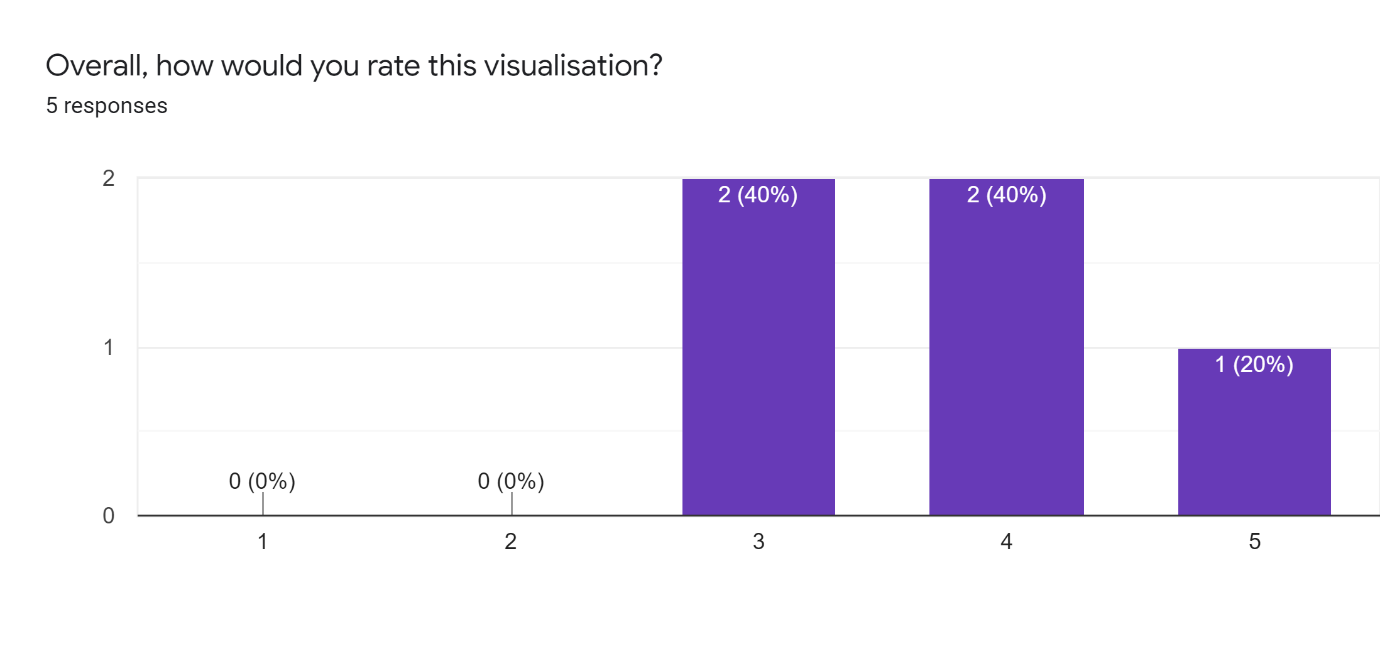


Figure : Rating of the Zebra system (Very bad 1-5 Very Good)

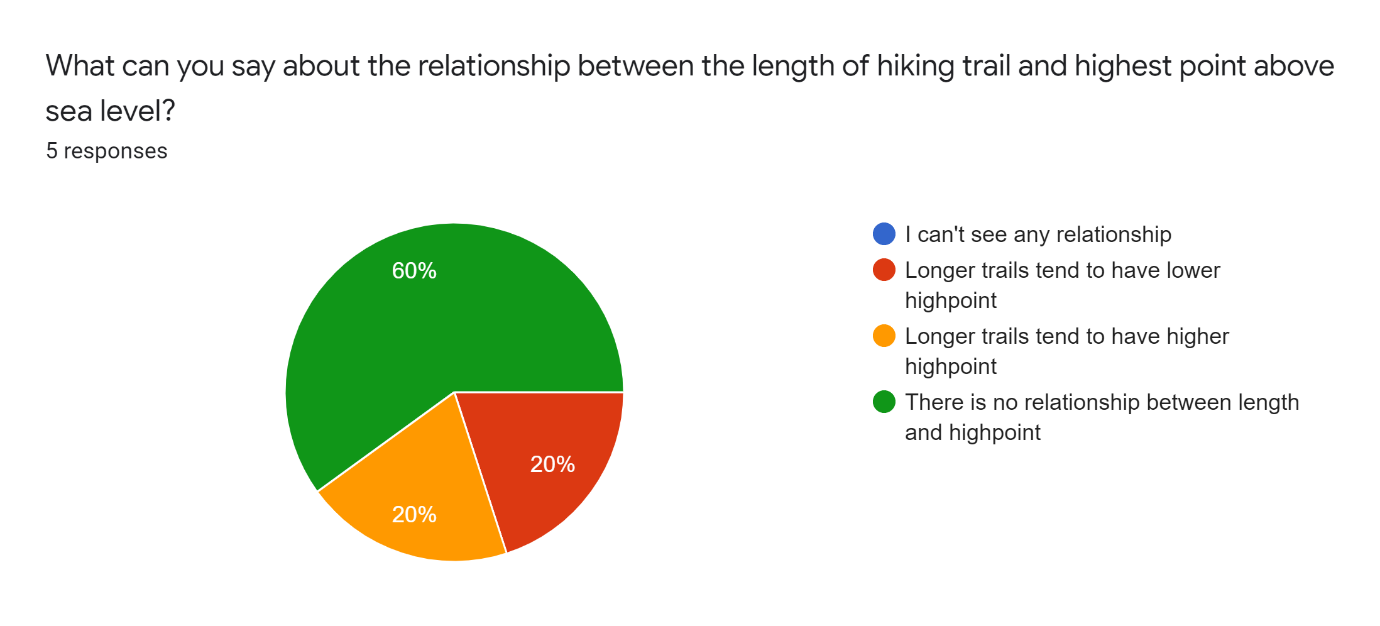


Figure : Relationship in the data perceived in Giraffe visualisation

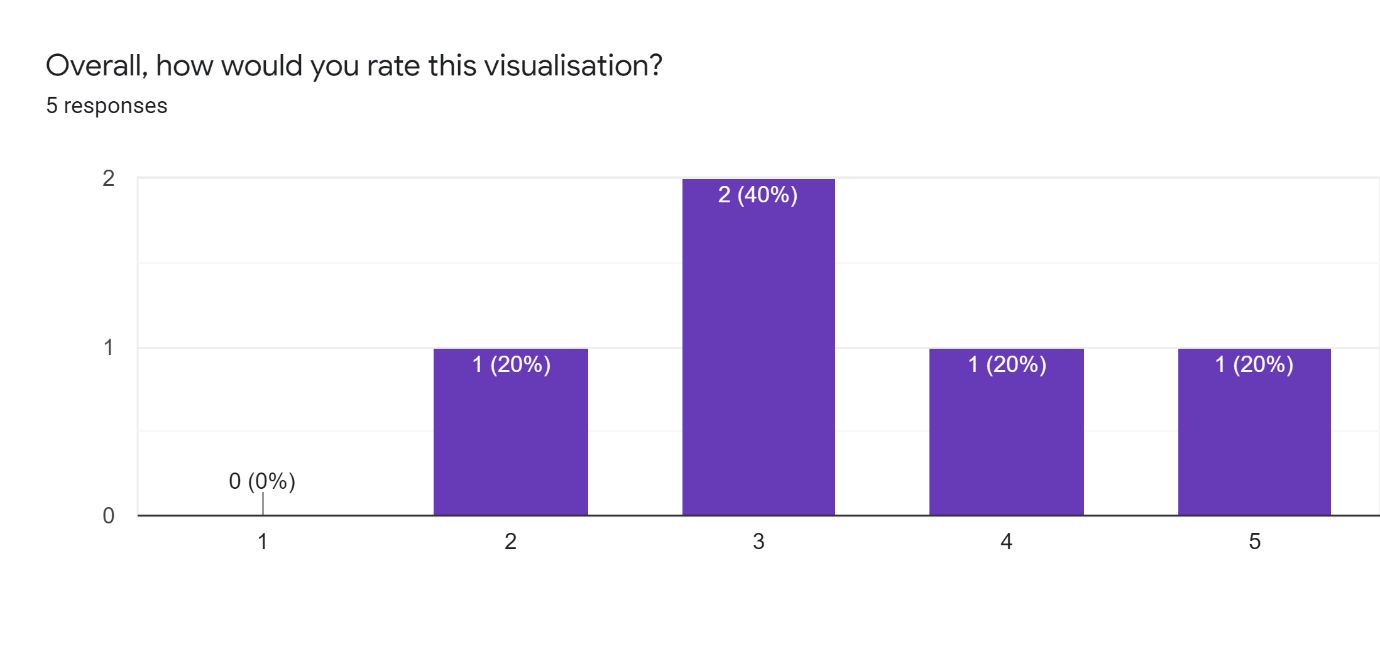


Figure : Rating of the Giraffe system (Very bad 1-5 Very Good)

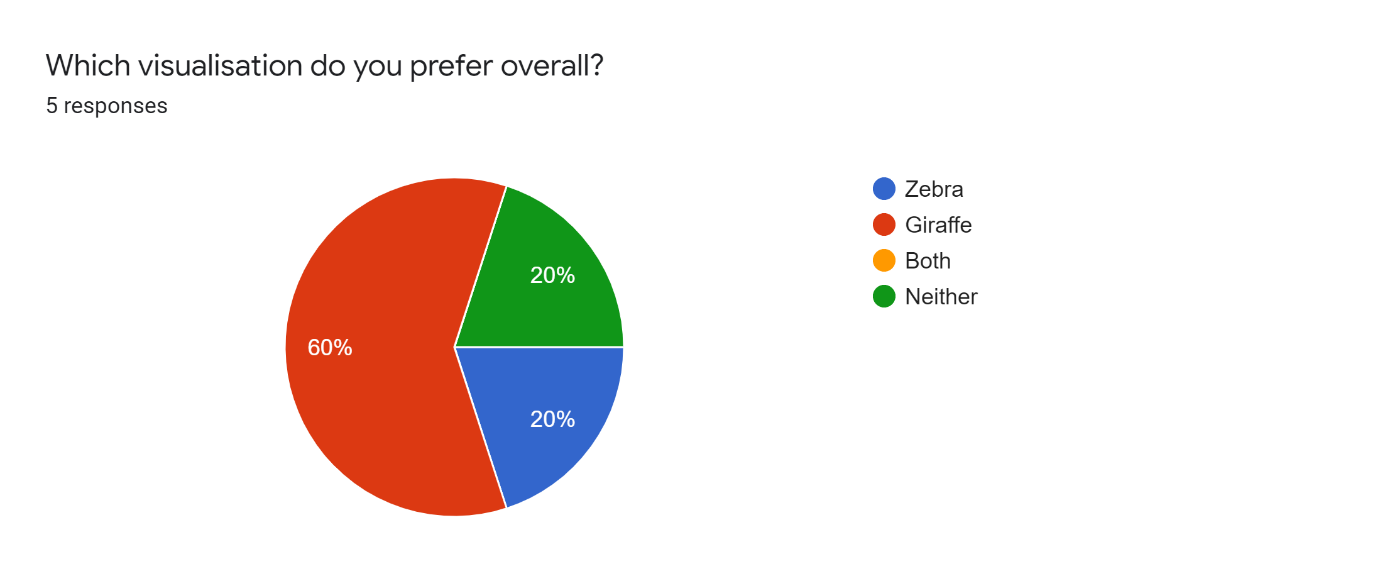


Figure : Evaluators preference between visualisations

# Future Work

[100 words]

# Appendix – Evaluation Data

Table : Participant Demographics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Participant | 1 | 2 | 3 | 4 | 5 |
| Age | 21 | 23 | 25 | 24 | 24 |
| Gender | Female | Female | Female | Male | Male |

Table : Evaluation of the Zebra System, correct answer indicated in bold

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Participant | 1 | 2 | 3 | 4 | 5 |
| Can you select the Mount Rainier Area? | Yes | Yes | Yes | Yes | Yes |
| Can you identify the longest trail of Mount Rainier Area? Please type its name: | Indian Bar Summerland Terrain | **Wonderland trail** | **Wonderland trail** | **Wonderland trail** | **Wonderland trail** |
| Can you identify the trail with the highest highpoint in Mount Rainier Area? Please type its name: | **Glacier Basin Loop** | **Glacier Basin Loop** | Mount Adam South Climb | **Glacier Basin Loop** | **Glacier Basin Loop** |
| What can you say about the relationship between the length of hiking trail and highest point above sea level? | There is no relationship between length and highpoint | There is no relationship between length and highpoint | There is no relationship between length and highpoint | There is no relationship between length and highpoint | There is no relationship between length and highpoint |
| Would you be able to choose a hiking trail to visit based on this visualisation? | Yes | Yes | Yes | Yes | Yes |
| Overall, how would you rate this visualisation? | 5 | 4 | 4 | 3 | 3 |
| What do you like or dislike about this visualisation? | I like that the layout is very straightforward | the units of measurement are clearly displayed, there are not too many data points - this is good | It requires a bit of concentration. If I'm going for a hike views would be the number one thing. Also if you are not used to using visualisation this can be confusing. | Very simple to glean information on length/ height from, but little else to go off of | Dots are too small and hard to correctly select with ease |

Table : Evaluation of the Giraffe system, correct answer indicated in bold

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Participant | 1 | 2 | 3 | 4 | 5 |
| Can you identify the longest trail? Please type its name: | **Wonderland Trail** | **Wonderland Trail** | Mount Adam South Climb | **Wonderland Trail** | Mount Adam South Climb |
| Can you identify the trail with the lowest highpoint? Please type its name: | **Mount Adams South Climb** | **Mount Adams South Climb** | North Coast Route | **Mount Adams South Climb** | No, there are 4 and the specific heights are not given |
| For the longest trail, are dogs allowed to visit? | **Dogs are not allowed** | **Dogs are not allowed** | Dogs are allowed on leash | **Dogs are not allowed** | **Dogs are not allowed** |
| What can you say about the relationship between the length of hiking trail and highest point above sea level? | Longer trails tend to have higher highpoint | There is no relationship between length and highpoint | Longer trails tend to have lower highpoint | There is no relationship between length and highpoint | There is no relationship between length and highpoint |
| Would you be able to choose a hiking trail to visit based on this visualisation? | Yes | Yes | Yes | Yes | Yes |
| Overall, how would you rate this visualisation? | 5 | 2 | 3 | 4 | 3 |
| What do you like or dislike about this visualisation? | The graph was clear and the shades of green were visually pleasing | It is not visually appealing, it is harder to make sense of it | Again too much work. | Much more info given, nothing sacrificed in simplicity | Axes are too small |

Table : Comparison and user preference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Participant | 1 | 2 | 3 | 4 | 5 |
| Which visualisation depicts hiking trail length and highpoint better? | Giraffe | Zebra | Giraffe | Giraffe | Zebra |
| Do you find features of the hiking trail relevant? | Yes | Yes | No | Yes | Yes |
| Which visualisation do you prefer overall? | Giraffe | Zebra | Giraffe | Giraffe | Neither |
| Do you have any suggestions on how the visualisations could be improved? | Nope | do not use giraffe visualisation, it confuses users | It's not very user friendly. The Zebra visualisation wouldn't work on my phone though |  | Make dimensions consistent. Increase size of text and icons so they can be easily seen |

# Bibliography

[1] Thomas Mock (2021). Tidy Tuesday: A weekly data project aimed at the R ecosystem. <https://github.com/rfordatascience/tidytuesday>.

[2] Elis Hughes, Patrick Ward (2020). Tidy Tuesday: Washington Hiking.   
<https://github.com/rfordatascience/tidytuesday/tree/master/data/2020/2020-11-24>

[3] *TEXT data type - IBM Documentation*. (n.d.). Retrieved July 21, 2021, from <https://www.ibm.com/docs/en/informix-servers/12.10?topic=types-text-data-type>

[4]