Design & Implementation of 2 Visualisation Systems for Hiking Trails Data

The Residuals

Table of Contents

[I. Video Demonstration 3](#_Toc78109354)

[1. Data Set Description 3](#_Toc78109355)

[2. Visualisation Objectives 3](#_Toc78109356)

[3. System Implementation 3](#_Toc78109357)

[4. Design Comparison 3](#_Toc78109358)

[5. User Evaluation 4](#_Toc78109359)

[6. Future Work 6](#_Toc78109360)

[A. Appendix – Evaluation Data 7](#_Toc78109361)

[Bibliography 10](#_Toc78109362)

# 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>

Alternatively, both systems and survey can be accessed [here](https://ananthp189.github.io/residuals/).

# 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 1: 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

The main idea was to develop two systems that could be used by visitors of the Washington Trails Association website for choosing hiking trails. The Zebra system is aimed at hikers interested in the length and highpoint of hikes, or relationship between these variables. The Giraffe system was inspired by tabular presentation of data. The hikers care about the rating, features or description of the hiking trail within the range of length and highpoint they are comfortable with. For both systems, length is on the X-axis and highpoint on the Y-axis. This was chosen to reflect natural geography of mountains.

The design decisions made in development of the Zebra and Giraffe systems were as follows:

* **Type of chart:** Zebra is a scatterplot of hike length against highpoint, where point represents a hike in the selected region. Giraffe is a heatmap of hike length against highpoint, where each square highlights the top-rated hike along with its features, in the highpoint and length grouped ranges. Both these charts allow the user to identify, compare, and browse the hikes.
* **User selection:** Zebra has a drop-down menu with different regions in Washington state as the options. This allows user to select region which will display the hikes (or points) within this region. Giraffe doesn’t have this feature, instead this system uses hikes from all the regions.
* **Data points shown:** Zebra displayed 15 top rated hikes from a selected region on the plot. Giraffe selects the top-rated hike depending on the binned highpoint and length ranges. Then the system displays the highest rated hike on the heatmap. Giraffe is choosing from the top rated 1000 hikes of Washington state. This system’s hikes meet the following conditions: to have a length less than or equal 100 and a rating greater than or equal to 3.5.
* **Binned axes:** Zebra’s highpoint and length axes have continuous scales, so the user can see the exact highpoint and length of the hikes. Giraffe’s highpoint and length axes have been cut-off by putting them into bins. Their highpoint and length axe’s bins are ordered by increasing 2000 feet and 10 miles, respectfully. The user can look up the top-rated hike by the binned the ranges.
* **Information upon hovering:** Both systems allow the user to hover over the points or squares and information about the hike will display over them. When hovering the points, Zebra only shows the name of the hike. Whereas the squares in Giraffe shows the name, rating and features of the trail.
* **Colour scheme for user rating:** Both systems have legends with allow the user to compare the ratings of the displayed hikes. Zebra has a gold and greencolour scheme to differentiate the display hikes ratings from each other. Giraffe has a gold and green colour scheme since some trails with lower rating were less visible using yellow and green. The white space in the Giraffe system means there are no hikes in those grouped ranges. Both colour schemes are inspired by the forest, to reflect an outdoor theme.

# User Evaluation

The summative evaluation was performed via an online survey after an initial formative evaluation by the team. A html page containing instructions and both visualisations were shared to participants along with a link to the survey. There were 5 participants in total, this included 3 females and 2 males in their early 20s. Tasks were designed to identify accuracy of interpretation and preferences of the users. Performance of users was not a design goal and hence not included in the evaluation.

Table 2 summarizes the results of the survey. Raw survey data can be viewed in Appendix A. Figures 1, 3 show overall rating participants gave to Zebra and Giraffe respectively. The median rating was 4 for Zebra and 3 for Giraffe. Figure 2 shows Giraffe led participants to discover more relationships in the data. Figure 4 summarizes the preferences of the users. 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.

Overall, participants had better accuracy of interpretation using the Zebra system. While Giraffe received lower median rating individually, in the comparative evaluation it was preferred by most participants.

Table 2: User Evaluation Summary

|  |  |  |  |
| --- | --- | --- | --- |
| System | Zebra | Giraffe | Comments |
| Able to filter region using dropdown? | 5/5 | NA |  |
| Able to identify longest trail? | 4/5 | 3/5 |  |
| Able to identify highest point? | 4/5 | 1/5 |  |
| Able to identify relationship between length and highpoint? | 5/5 | 5/5 | No relation observed by all |
| Able to identify features of longest trail? | NA | 4/5 |  |
| Able to shortlist a hike? | 5/5 | 5/5 |  |
| Preference | 1/5 | 3/5 | 1 participant preferred neither. |
| Positives | 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 | colour scheme, the simplicity of the graph, with more information given |  |
| Negatives | points are too small and hard to select,  little information beyond what is shown and that it requires concentration especially to new users | axes are too small; graph was harder to make sense of and not visually appealing |  |
| Median Rating (1-5) | 4 | 3 |  |

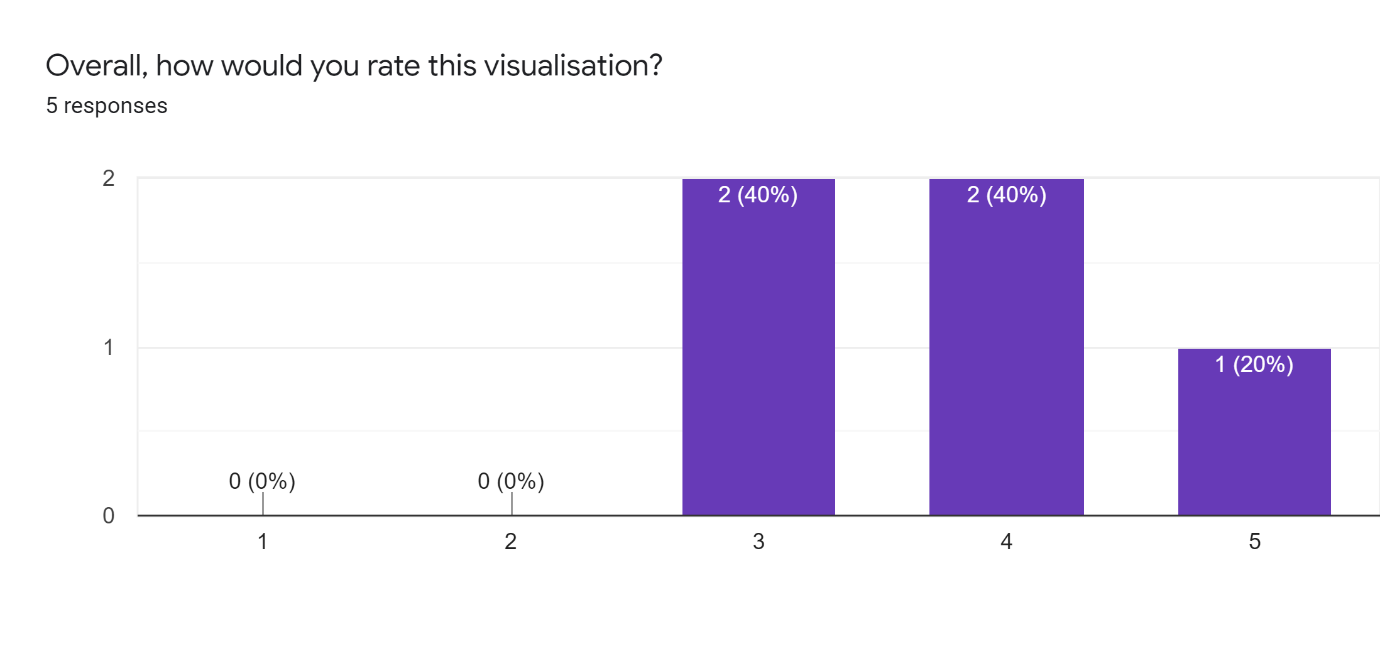


Figure 1: Rating of the Zebra system (Very bad 1-5 Very Good)

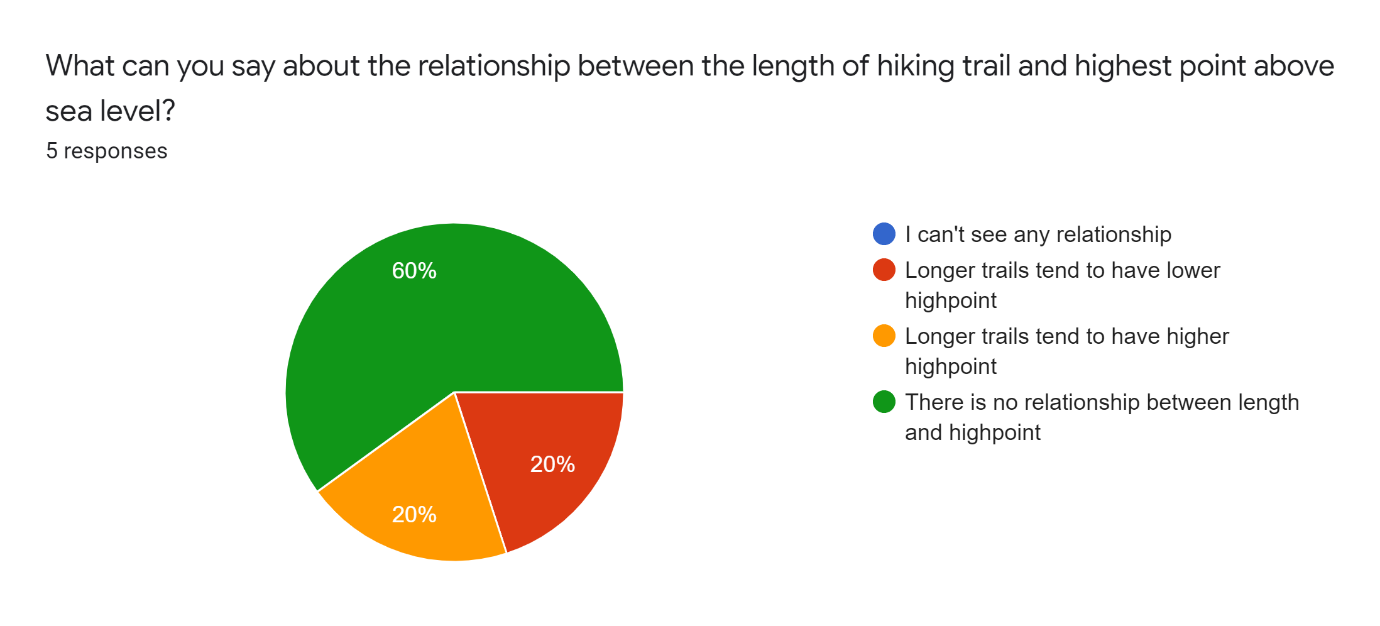


Figure 2: Relationship in the data perceived in Giraffe visualisation

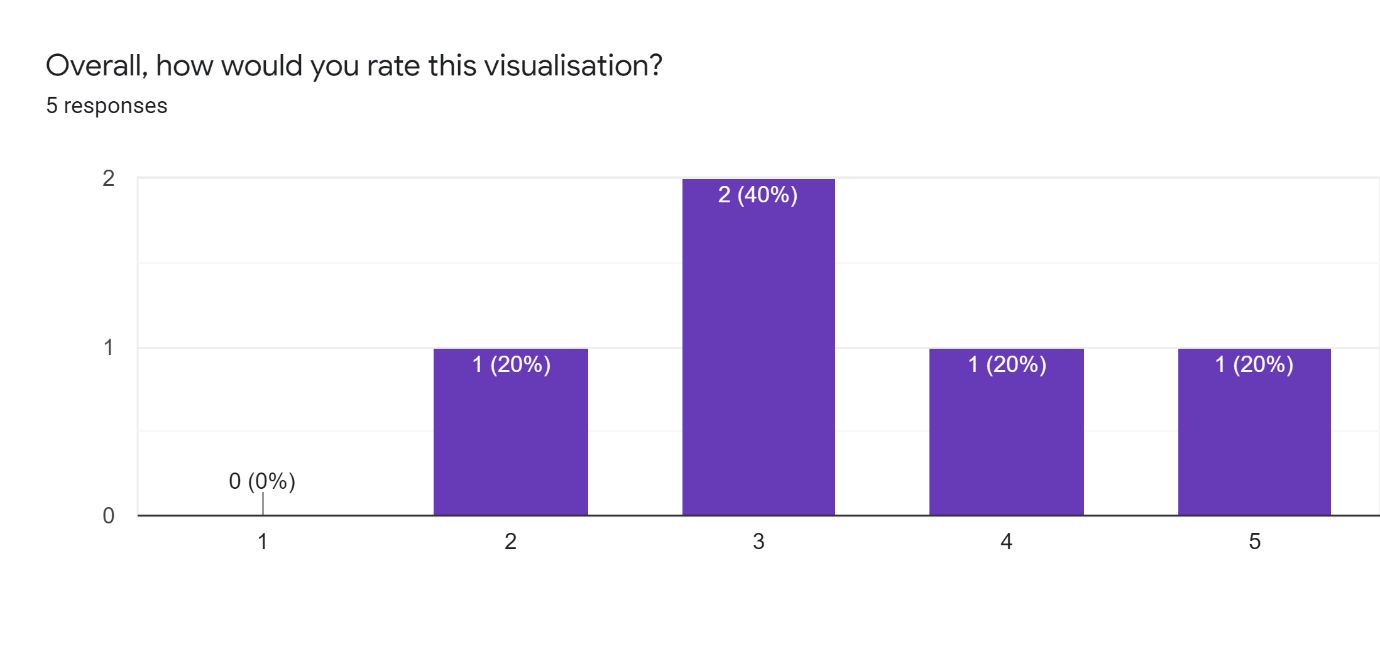


Figure 3: Rating of the Giraffe system (Very bad 1-5 Very Good)

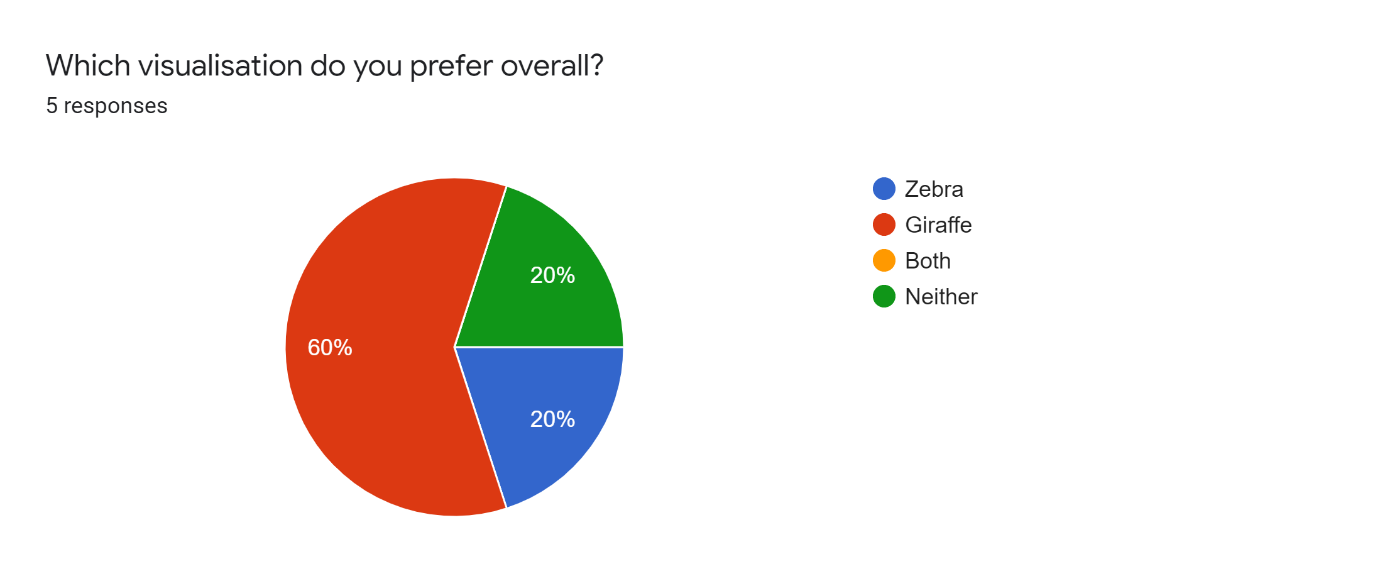


Figure 4: Evaluators preference between visualisations

# Future Work

For the Zebra system, scatterplot points and text size will be made bigger. Hovering over the trail will additionally show user rating and features, which is consistent with Giraffe. For Giraffe, selecting region can be incorporated. The Y-axis will be adjusted in Giraffe. For both systems, visualisations showing length and gain can be added. Description of the trail could be added to the information box.

# Appendix – Evaluation Data

Table 3: Participant Demographics

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

Table 4: 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 5: 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 6: 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>