

A4 REPORT

COURSE: VIS WS '18 DATE: 13.12.2018

Name : Robert Ernstbrunner
Matnr.: 01403753

1 Introduction

The ornithology student *Mitch Vogel* (yes, he's a birdwatcher named *Vogel*) needs our help in visualizing his data in order to advance in his investigations. There is an observed decline in the nesting *Rose-crested Blue Pipit* in the *Boonsong Lekagul Nature Preserve* and Mitch thinks the traffic going through the preserve might have to do with it. He reckons that traffic noise might overshadow mating calls or that invading campers might scare away the birds from their habitat areas.

The data was collected by park rangers who work as caretakers of the nature preserve. They gave Mitch additional explanations about the data together with a map.

2 Data, Users, Tasks

The park rangers monitor the traffic through the preserve with the help of sensors. Tickets are assigned to vehicles at the entry gates that get scanned by these sensors when the vehicles pass by them.

- In summary there are seven different vehicle categories:
 1. *Cars* and *motorcycles* count as one type.
 2. There are three types of *trucks* (differentiated by their number of axles).
 3. There are two types of *buses*.
 4. The seventh type consists of *park service trucks* that are used by the rangers who have access to all areas.
- There are five different sensor types:
 1. Vehicles that enter or leave the preserve are tracked at *entrance gates*.
 2. *General gates* act as control points to monitor the traffic flow.
 3. Special gates, which are simply referred to as *gates*, prevent general traffic from entering restricted paths. These gates can only be opened by rangers.
 4. *Ranger-stops* represent the working areas for rangers. Some are cut off from general traffic and some aren't. That's why some of their sensors also collect general traffic information while others don't.
 5. Finally *camping* sensors monitor visitors, that enter or exit a camping ground.

- Additional information about the dataset:
 1. Traffic either passes through the preserve, stays as day campers or stays as extended campers.
 2. Rangers reside at the southeast ranger-stop when they are not working.
 3. The speed limit is 25 mph.
 4. Summertime is not included in the sensor measurements.
 5. Traffic going southward from entrance gates *three* and *four* are not represented on the map and are not tracked.
- Data snippet:

```
Timestamp,car-id,car-type,gate-name
2015-05-01 00:15:13,20151501121513-39,2,entrance4
2015-05-01 00:32:47,20151501121513-39,2,entrance2
2015-05-01 01:12:42,20151201011242-330,5,entrance0
```

The data set consists of the following parameters:

1. *Timestamp* (**quantitative**) contains the date and time the sensor was triggered.
 2. Entering vehicles are assigned with a *car-ID* (**nominal**) at the *entry gates*.
 3. The *Car-type* (**nominal**) specifies a vehicle according to their type and number of axles. A park vehicle gets additionally marked with a 'P'.
 4. The *Gate-name* (**nominal**) is the name of the sensors themselves. The same names can also be found on the map.
- The preserve:
The *preserve* is open to the public 24 hours a day all year long. It is visited by naturalists of all ages, photographers and explorers. It is used for hiking, camping, nature study, *birdwatching*, photography and cycling. The preserve can be entered through five entrances by paying an access fee. According to the map, it comprises nine camping areas.
 - The included map is presented in [Figure 1](#). The map consists of a black background with white roads and type-colored gate labels. It may be used as a template by the flow map visualization ([design 2](#)).

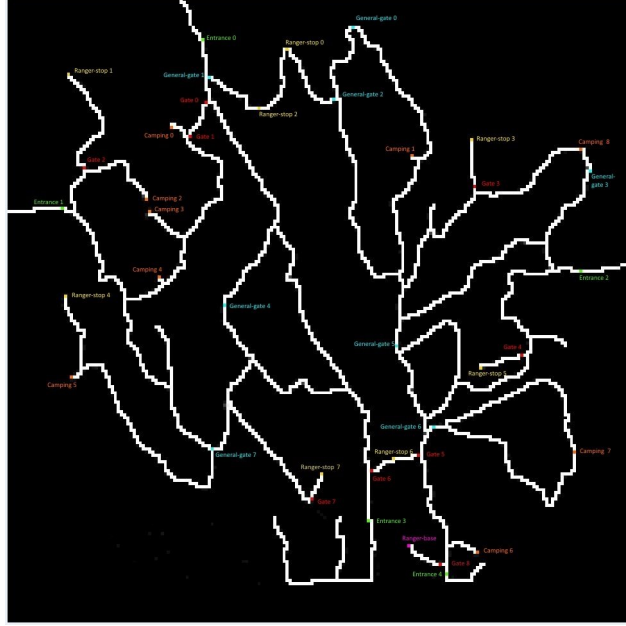


Figure 1: the rangers map

3 Designs

Design 1

Visualizing behavior inside the preserve. For my first design (Figure 2) I concluded that *Mitch* might want to gain insight on the traffic behavior inside the preserve. Therefore, I tried to find possible correlations between the time spent at different sensors over a single day. The most significant separator between sensor types would be the one between camp sensors and the rest. Having a closer look at relations between these two sensor types might answer the question if most of the visitors are serious (longterm) campers or just visitors/explorers spending their time mostly elsewhere. This might answer *Mitch's* question if campers are mainly responsible for the decline in birds. Therefore, I drew a scatterplot in the left upper corner. The horizontal axis shows the duration spent at camps and the vertical axis shows the duration spent elsewhere. The dots represent the vehicles and since there are only seven categories, I decided to distinguish them by color.

Just examining the relation between camp sensors and the rest might lead to fast conclusions but it might not be enough to gain deeper insight and understanding of the data. Therefore, I sketched a scatterplot matrix (*SPLOM*) right next to the scatterplot to reveal more details about the time spent at each sensor type and their correlation to one another. I also included the overall duration to (hopefully) make it even more insightful (it might be a good idea to limit the

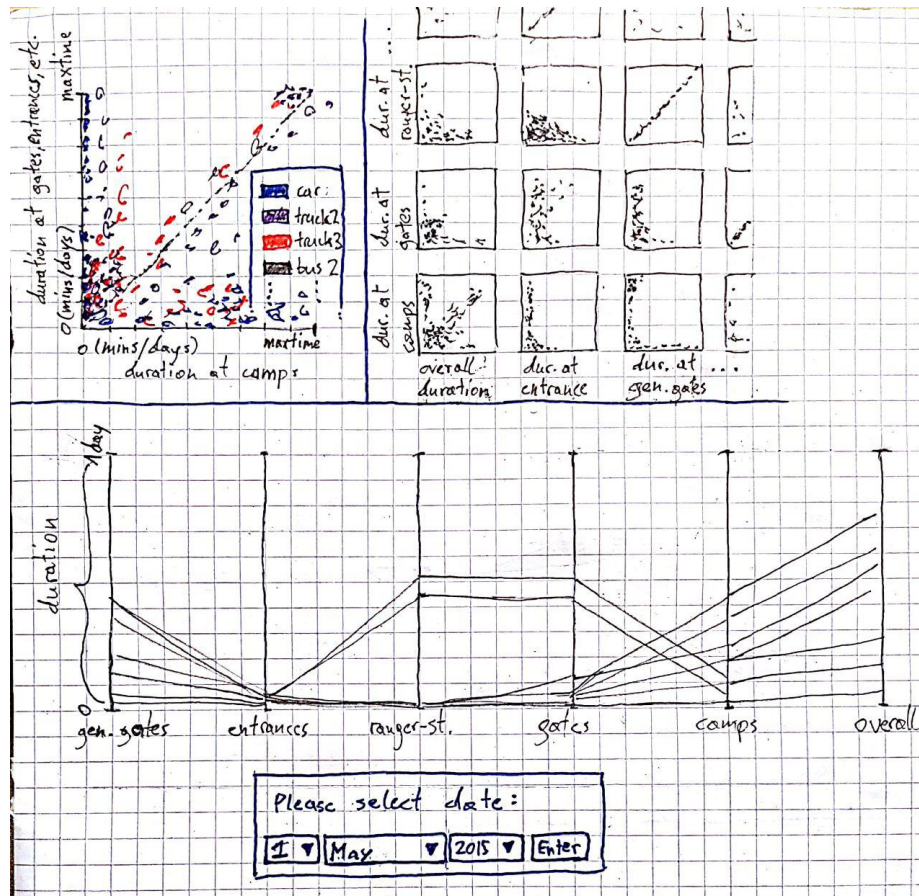


Figure 2: Design 1

overall duration for a vehicle to a single day in order to preserve good visibility). The *SPLOM* requires a table with ordinal data as input. Therefore, a table with duration at different sensor types has to be computed from the data. I was thinking about eventually removing the redundant upper triangle together with the diagonal, in order to save space (this is commonly done [Munzner, 2015]), but then I realized that this might inhibit readability for *Mitch* who might have never seen a *SPLOM* before. I think someone in the *VIS* course once said *having to explain a visualization is mostly a sign for a bad visualization*. Also, I was thinking that the *data ink ratio* is not affected as long as the (redundant and transposed) data in the upper triangle is justified. It could be argued that *Mitch* might not understand the *SPLOM* either way. That’s one of the disadvantages of *SPLOMS*. They might eventually reveal deeper information but are initially harder to read and understand.

Parallel coordinates (*PC*), on the other hand, are more accessible, at least to me. That’s why I also included them in the design below the scatterplot and *SPLOM*. *PC*, like the *SPLOM*, also show correlations, but through different kinds of visual patterns. While a dot-pattern pointing upward / downward means positive / negative correlation in the *SPLOM*, having parallel / mostly intersecting lines between axes means positive / negative correlation in *PC* [Munzner, 2015]. In both cases understanding and finding correlations is not intuitive and must be learned beforehand. My *PC* consist of duration axes for the five different sensor types together with the overall duration axis. Interaction would, besides selecting the day, then include reordering these axes for the *PC*.

Design 2

Visualizing traffic flow. For my second design (Figure 3) I wanted to show what roads are most frequently used overall, because this might correlate with the area where the birds decline the most.

One could argue, that the map from the rangers (as a visualization on its own) already contributes valuable information by revealing the camp- and road locations. Therefore, the map could be used to draw reasonable conclusions right away. Since the map can be represented by a planar graph, I chose to visualize this with a *flow map*. The sensors are represented by circle-shaped vertices. The vertices should have similar spacial placing like the sensor dots depicted on the ranger map. The paths of the vehicles are presented by merged edges which should also be approximated to the spatial layout of the ranger map. The edges grow in thickness according to the traffic intensity on the paths. The vertices also grow (in a continuous manner) in size to match the thickness of the edges (this is a questionable design choice and in hindsight, I should have drawn them with equal size).

The total amount of vehicles is presented in an additional barchart and lets *Mitch* figure out what vehicle type is most common in the preserve.

Selecting a bar highlights the merged edges for the selected type and, while still visible, puts the unselected edges from the merged edge set in the background. This would then reveal the ratio between the selected type and the unselected

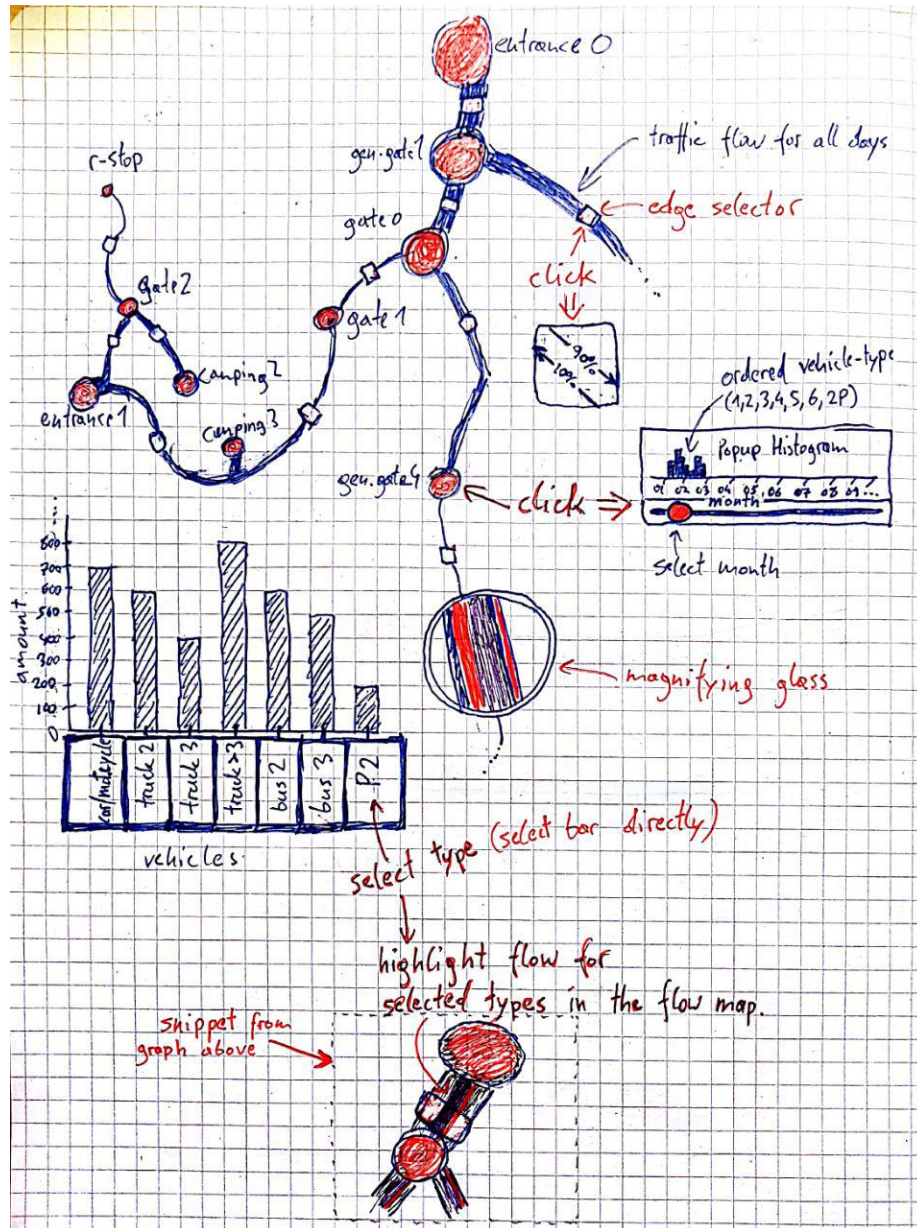


Figure 3: Design 2

types directly on the paths.

A drawback might be that the graph, in reality, is a directed graph, but is presented as undirected, leaving out valuable information in what directions the vehicles were going. Therefore, I implemented square-shaped edge selectors on the merged edges that, when clicked, will yield a popup. The popup then contains two parallel bi-directed arrows, with the same slope as the merged edges at the edge selector location, that present percentages of the directions that all the vehicles on that path were heading in.

Clicking on a vertex also yields a popup. This popup contains a timeline histogram with a slider to select the month. The histogram shows the total amount of vehicle types that were passing the sensor in the selected month. This might help *Mitch* evaluate the traffic for a specific time domain on a specific path.

Additionally, a magnifying glass might also help revealing further traffic information, especially from thin edges. I prefer a magnifying glass over a fisheye lens, because the distortion obtained by a fisheye lens heavily impairs distance / length judgments and thus simply looks and feels horrible to me. I was playing with the idea, to show the histogram of a vertex when hovering the magnifying lens over said vertex. I figured the lens could eventually still be too small, also I wasn't brave enough to make a (much beloved) pie chart inside a magnified vertex. Therefore, the magnifying glass would show nothing new when dragged over a vertex.

Design 3

Visualizing traffic density (and traffic movement). For my last design (Figure 4) I chose to work mainly with stacked visualization types to visualize traffic density. The stream graph shows the amount of vehicles that are in the preserve over one day. Since the amount of vehicles is *quantitative* data, and the vehicles themselves are *nominal* and few (i.e. *colorable*) and time is an *ordinal* attribute, this type of visualization was a perfect choice in my opinion[Munzner, 2015]. The day can also be selected with the date selector in the bottom right corner of the design.

Below the stream graph I wanted to show movement between the different sensor types. Therefore, I drew histograms at different points in time that show the amount of vehicles currently present at each sensor type. This could have probably been animated as well, but I consider my approach to be superior over animation because the change in the five histogram bars would be too much to comprehend in an animation. Generally, in such cases, the amount of frames should always be reduced, i.e. convert from animation to side-by-side views. In total, there are twelve views distributed over two rows. It was of very importance to me that the views are row-aligned in order to be closer together. This should reduce cognitive load and there is only one line jump were *Mitch* has to rely on his memory instead[Munzner, 2015]. Everyone understanding the english language tends to read from left to right. This habit should also adapt to the side-by-side view.

Another thing that might interest *Mitch* is the duration that vehicles spend at

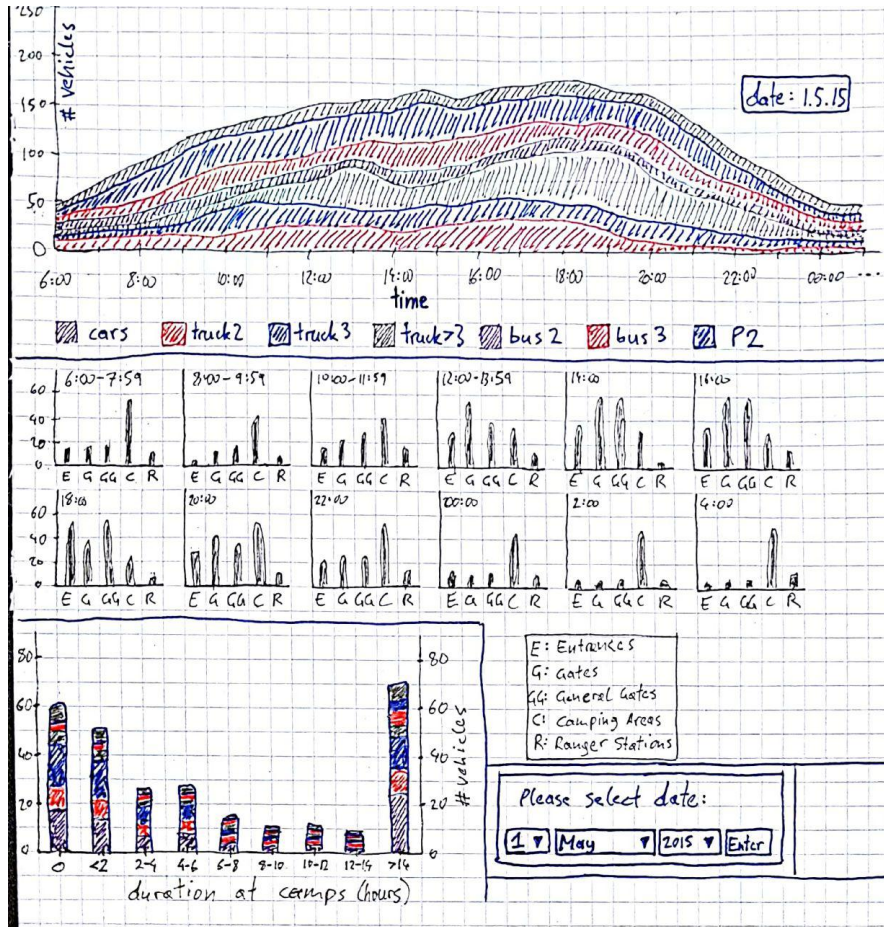


Figure 4: Design 3

camping areas. Since the vehicle types are nominal and colorable, the duration intervals are nominal and the amount of vehicles is quantitative, it made sense to create a stacked bar chart here[Munzner, 2015].

4 Comparison

I reckon design 2 is probably a bit too complicated for *Mitch*, also because it contains by far the most interaction from all three prototypes. Therefore, there is a high chance that he could overlook or miss some of it. That's why e.g. the *edge selector* must be designed in a way, that *Mitch* wants to click on it intuitively. I justified the *vertex selector* earlier, stating that this might help *Mitch* gain insight in the traffic over a month for a specific path. I felt the need to justify this because this gimmick somehow feels a bit tucked on. If I had to

rethink my design, I would probably re-imagine the *vertex selector* first. However, I consider the *flow map* itself extremely valuable, because it works good for visualizing the movements of objects from one location to another and it reduces visual clutter by merging all the edges. Therefore, it has high potential in conveying important information about the spacial traffic distribution. I chose visualization types like *flow maps*, *bar charts* and *histograms* that work for large amounts of data. (Also this is so trivial I can't believe I'm even writing this, but the bar chart works for nominal data (like *vehicle type*) and line charts for ordinal (like *time*), therefore the *bar chart*.)

I consider [design 1](#) to be the most developed one because the data is transformed into something else and then visualized by a *SPLOM* and *PC*. I chose the time range to be one day because this should reduce the amount of scatterpoints to a few hundred. This kind of time filtering is needed in order to visualize the data properly. Too many points would probably yield in a *PC*-hairball, revealing almost no information at all. A drawback might be that the *SPLOM* might not show any correlation at all, and then I would have to tackle the problem differently.

It is hard to compare this design to the previous one, because the tasks differ. Also one is an overview and the other one provides detailed visualization.

The [third design](#) kind of merges the two previous tasks. It looks mainly at traffic density (behavior) via the *stream graph*, *stacked bar chart* and *gate histograms* and a little bit at traffic flow by examining the gates at different steps in time via the side-by-side view of the *gate histograms*. This design would also work as an overview, since all visualization types are capable to display large amounts of data.

5 Summary

In order to create my visualizations I needed estimates about the amount of data entries with respect to different attribute constraints. Luckily, the data was already there and I could lookup different things by pre-filtering in *Excel*. This helped in choosing time ranges for the detail design views.

I noticed that only arrivals at gates / camps are tracked, but no departures. Therefore, we could come up with the assumption, that campers could also camp somewhere along the way before the next gate. However, in the Lekagul Preserve description file, it clearly states, that camping and cycling are only allowed in designated areas. So if the rangers do their job properly, (which we could check by visualization) it is reasonable to strongly assume the opposite.

Also, there is no information about the birds themselves. This fact limited the visualization space by narrowing the visualization tasks to the traffic only.

The third design addresses both tasks (traffic flow and traffic behavior) and would also work not only for a single day, but also as an overview, since all visualization types are capable to display large amounts of data as well. Having this kind of flexibility and a high chance of displaying valuable information (compared to the first design) I consider this to be my best prototype overall.

5.1 Realization in Tableau

I imagine the *flow map* would be the hardest to do in *Tableau*, because it would have to be drawn by hand as far as I know[Munzner, 2015]. Also one would have to calculate the edges and implement all the interactivity, which is a bit of an overkill, I think. The first design would require the data to be transformed into a duration table fist, which is manageable, but Tableau doesn't support *PC* at the moment, which is unfortunate. The easiest design to realize would probably be the third prototype. Tableau can easily handle stream graphs, histograms and stacked barplots.

References

- T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015. ISBN 9781498759717. URL <https://books.google.de/books?id=NfkYCwAAQBAJ>.