# Final hand-in for visualization

Lasse Ahlbech Madsen, Simon Maibom

**Abstract**— Best abstract in the world, it so great you just won't be able to handle it. It does something, no one has ever done before. It utterly destroys its job as an abstract. The world will not be the same when you are done reading this amazing abastract. Read it and weep bitch.

**Index Terms**—Settlers, Catan, Visualization

## 1 Introduction

We have decided to work with the boardgame Settlers of Catan. The idea is to visualize the effects of starting choices on the outcome of the game. It is a game about managing odds and hedging your bets, so frequent players will be interested in knowing what types of starting locations that gives a high probability of winning.

### 1.1 About the game

Settlers of Catan is a game where you and a number of other players settle an island. During the course of the game you acquire resources, build improvements and acquire points. The first to 10 points wins the game.

For our project it is important to know that a settlement is located next to 3 resource tiles with a number tile on top. Whenever the dice outcome is the same as a number of a tile all players next to it receive a resource token. Settlements can also be located next to the sea where there are harbors that can be used to trade resources.

## 2 Related work

There have not been much work done on the topic, we have not been able to find actual visualizations of the game, but we did find one analysis by Peter Keep [1]. He attempts to use probabilities to estimate the worth of resources and getting a gist of an optimal strategy. He never refers to his dataset, but merely states things as known. He makes a good argument for his approach to find the value of resources, but in his conclusions he fails to take rarity of resources fully into account.


- *Lasse Ahlbech Madsen, E-mail: xsc606@alumni.ku.dk*
- *Simon Maibom: E-mail: xvm226@alumni.ku.dk*


*Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx*

## 3 Data and Task Abstractions

### 3.1 Domain Situation

The target audience of the project is intermediate Settlers of Catan players, who wants to get better at the game by getting a deeper understanding of how to place the two starting settlements in order to maximize the chance to win.

Relevant questions these players might ask would either be very concrete in-the-moment questions or generally browsing for good combinations and things to avoid. We imagine most players would be interested in just browsing the data as well.

- "What is the optimal settlement location?"
- "Is it feasible to settle without access to all resources?"
- "Does settling next to a desert always cause a loss?"
- "If my first settlement has wood and ore what resource should I go for?"

### 3.2 Data and Task Abstractions

The dataset [2] consists of data for 50 4-player games. Each game has 4 lines that consist of starting position, points gained, placements of starting towns, total resource gains and losses from production, robber cards and trade.

A snapshot of the dataset can be seen in appendix 1. We have mostly used the data from the settlements and the points. The types of the attributes can be seen in table 1. It is worth nothing that the dices on the resources are categorical because comparisons of them in this context is meaningless.

| Data | Type |
|---|---|
| GameNum | Ratio |
| Player | Categorical |
| Points | Ratio |
| Me | Categorical |
| Dice throws | Categorical |
| Settlement:res | Categorical |
| Settlement:dice | Categorical |
| Gains | Ratio |
| Losses | Ratio |

Fig. 1. Attribute types

To win a game of Settlers you need 10 points and there will never be two players with 10 or more points at the same time. However it is possible to finish the game with 11 or 12 points, but this skews the data without being particularly interesting when determining a winner, so we changed all values above 10 to just 10.

Generally points are the main indicator of success in the game, so whenever we need to analyze the quality of combinations we will measure it against points and occasionally win-rate.

We did not consider gains and losses all that interesting. They are mostly based on chance that is dependent on your settlements. We could show them, but it would merely be a correlation between high gains and winning. The experience Settler player would not be particularly interested in seeing this.

## 4 Design

Based on our data and task abstractions it would seem that the most important part is the ability to view probabilities and to search the data for exact situations. There are too many possible combinations for pie charts and the probabilities do not relate to each other, so we stuck with simple bar charts for the main part of our visualization.

What was more important was the filtering of data and we want to ensure that any scenario can be created. We imagined doing this just with checkboxes, but we modified this for implementation.

### 4.1 Illustration

LAV EN GOD SKITSE

The sketch shows on the left side a bunch of filters that you apply, once the user have applied the filters on the right side will show a bunch of graphs of game statistics. These graphs include how many points gained relative to which number or resource a player had next to the starting settlements and other relevant statistics. The filters that the user should be able to apply, includes which player position, how many of a resource the start settlements border and whether they are next to a desert.

## 5 Implementation

We implemented the visualization using the D3-library. We did some preprocessing in python

## 6 Scenarios of use

### 6.1 Scenario 1

If my first settlement has two lumber and one ore what resources should I go for?

A user is in the middle of a game and has placed his first settlement next to two lumber tiles and one ore tile. He wants to know what he should look for when placing his second settlement.

He uses the drop downs menu and chooses settlements with at least 2 lumber and 1 ore. The resulting graph is still rather large, so he tries selecting ”group duplicate resources”. From this he is able to gather that sheep and wheat might just be the things to go for. Finally he scrolls to the second graph and chooses to show ”resource values based on second settlement”.

Sheep has the highest value, but he only wants to compare to people who have actually done well, so he sets minimum amount of points to 9. Sheep and Clay look equally strong. He checks the dice values and decides that while he can go for either or both, the dice value of the resource should be more important than which one he chooses.

### 6.2 Scenario 2

”Should the desert be avoided at all cost?”

A user wants to know if it is feasible to place a city next to desert if the other tiles are good. He ticks ”next to desert” and sees 8 bars of data. They are fairly spread, but he notices that some have won with it. He scrolls to the second graph and sees that ore and wheat are the most valuable resources in desert starts. He tweaks the graph and the requires data to be of players with more points. He notices that as he increases the number of required points wheat drops in value and ore remains the most important resource. He also notices that your tiles needs to have good dices on them to be valuable.

He concludes that good ore tiles can justify starting next to a desert.

## 7 Discussion and Future Work

Our visualization is not all we thought it would be. We had hoped that during our implementation, through working with the data, we would come up with more interesting visualization options, alas we are still using bar charts. However the charts do a good job with data exploration, the filters are pretty extensive and our second chart makes it pretty clear what you want to settle. Our project is a statistical project and simple bar charts do well in this field. Another issue is that our implementation does some statistical work, so the results would have been more interesting if we instead of 200 data points had 20000. We have tried to make up for the lack of data, by always writing the amount of data points, so the user will know to take it with a grain of salt if there is just 1 case.

We would have liked to spend more time on the design step of the framework, but everything got rushed so we had to prioritize getting started over getting the design right. While aesthetics is not the priority of this course we would have liked to make it all a little more pretty. The default colors in the second graph are particularly nasty to behold.

We never came up with a better way to show the resource combinations on the x-axis than their starting letters. We tried to do it with icons or colors, but it just became a mess, we also tried coloring the individual bars dependent on the resources involved, but if anything, it made it even worse. The initial graph is hard to read without filters applied, but this was never the main intent of the it. We want the user to filter down the data to something smaller, to explore the data by creating scenarios.

All in all our implementation is not all we had hoped it would be, but we feel it does a decent job at visualizing the data by doing the statistical work and allowing the user to explore it.

## Acknowledgments

The authors wish to thank the Kaggle user ”Lumin” for sharing his game stats

## References

[1] P. Keep. Settlers of catan analysis. Wordpress, February 2011.
[2] ”Lumin”. My settlers of catan games. Kaggle, August 2016.

## A Appendix 1: Dataset

Our dataset consists of 50 games each with 4 players. This makes for 200 lines of data. In tabel 2 a condensed version of the data can be seen. There are columns for all dice throws between 2 and 12.

| GameNum | Player | Points | Me | 2 | ... | 12 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 5 | | 1 | ... | 1 | | | | | |
| 1 | 2 | 9 | 1 | 1 | ... | 1 | | | | | |
| 1 | 3 | 10 | | 1 | ... | 1 | | | | | |
| 1 | 4 | 5 | | 1 | ... | 1 | | | | | |
| Set1 | | | | | | Set2 | | | | | |
| 6 | L | 3 | C | 11 | C | 9 | L | 10 | W | 11 | O |
| 5 | W | 8 | O | 10 | W | 4 | L | 5 | S | 11 | O |
| 5 | S | 6 | S | 12 | W | 8 | O | 4 | S | 3 | C |
| 6 | O | 9 | L | 3 | L | 4 | L | 8 | L | 10 | S |
| Production | TradeGain | RobberGain | TotalGain | TradeLoss | RobberLoss | Tribute | totalLoss | totalAvailable | | | |
| 38 | 5 | 2 | 45 | 10 | 2 | 4 | 16 | 29 | | | |
| 48 | 8 | 6 | 62 | 11 | 1 | 8 | 20 | 42 | | | |
| 44 | 14 | 9 | 67 | 24 | 4 | 0 | 28 | 39 | | | |
| 42 | 12 | 0 | 54 | 24 | 6 | 0 | 30 | 24 | | | |

Fig. 2. Dataset snapshot