Project Book

Team Tesla

Consumers Report for Electric Car Sales, Reliability, and its Contribution for the Ecosystem -- CO2

HARVARD
Extension School

Course of Studies:	CS 171 - Visualization
Supervisor:	Andrew Reece
Team Members:	Byron Bahan Enrico Mund Tim Hagmann
Prototype Website:	http://goo.gl/3f5Uf1

MOTIVATION & ABSTRACT

The U.S. emits more greenhouse gas emissions per capita than any other nation. Approximately 33% of the CO2 emission of the United States come from the transportation system. Moreover, a big part of the US growth in CO2 emissions since 1990 comes from the increase in CO2 emission out of transportation. Yet, in the last couple of years a new type of car has started to appear on US roads that is being hailed as the solution to the above problem, the electric car.

The project takes a deep dive into electric vehicle usage in the US. It is going to tell the story for a reader who could be a possible buyer of an electric vehicle in the near future. That mean it gives a general overview about electric vehicle technologies and how the environment can benefit from using it. Consequently that means guiding the reader through model overviews / market shares / sales statistics towards important usage questions like service and charging network.

KEYWORDS

Binary Decision Trees, C5.0, CART, Classification and Regression Trees, CO2, D3, Data Visualization, Electric cars, Infographic, R, Market share, Sales Statistics, Tesla, Treemap

TEAM EXPECTATION AGREEMENT

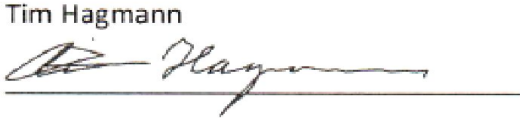
It will be the aim of our team to successfully complete the final project of CS 171 with the maximum of grade performance and to ensure the understanding of all tenets, facts, and errata connected with the study of the final project for all group members. To this end, the following terms and conditions will be followed.

Terms and Conditions.

1. Under any and all circumstances, I will get what (team) work I am allotted to do, done and turned in on time.
2. If I am sick and unable to make it on the date a team assignment is due, I will contact my team members to make other arrangements to get my work turned in on time. Should an emergency arise that prevents me from attending a team meeting, I will notify my fellow team members immediately.
3. The team will schedule weekly meetings.
4. I will do my share of the team work, there will never be an occasion where one team member does all of the work nor will there be a time when a member does none of the work.
5. I will do everything in my capabilities to help my fellow team members understand each and every concept and problem and I will not hesitate to ask my fellow team members for help. I will communicate with my fellow team members about any concerns I have with our group work. I will promptly report any team functioning problems to the instructor.

Acceptance: Each of the below signed team members agrees to abide by the terms and conditions outlined herein. Breach of this contract will result in a verbal warning the first and second offences. Third offence violation will result in dismissal from the team.

Tim Hagmann



Byron Bahan

Enrico Murid

1 Objectives and Goals

Our project will give a detailed overview of the current state of electric cars in the US. Our goal is to increase the reader's knowledge about this topic and to guide him through the many aspects. We will try to help the reader to form an opinion about this new technology.

Questions the project would like to answer:

Putting electric cars into context of the current environmental debate. At what magnitude do the individual states emit carbon dioxide and how much of this related to transportation. Can this be linked to usage of electric cars?

What is the economic situation of electric cars, do they sell? How many manufacturers offer electric cars in the US and how is the model variety and price range?

The electric car is a relatively new technology that entered the consumer market only recently, how is the technical reliability, are the users satisfied ?

Is it convenient to use an electric car ? Where can the consumer get service and where can he recharge his vehicle ?

2 Tasks

Using a choropleth map the reader will be able to select the displayed data and year to answer questions about spatial and temporal correlations between CO2 emissions and usage of electric cars.

Sorting a treemap by sales numbers will answer the questions about the economic situation of electric cars. The reader will be able to navigate through the various models of electric cars and open individual branches to retrieve detailed information about varieties and prices.

The question about reliability and customer satisfaction will be answered by selecting data and generating a respective chart (line, bar, pie). Zooming into an overview map will give details about service and charging stations for electric vehicles.

Feature List:

Visualisation 1 : Renewable energies for electric cars (choropleth map)

- Filter for specific energy type renewable / conventional
- Hover over each state to display state specific energy mix

This visualization should allow the reader to get an exact overview of the way how electric energy is generated in his state and how the use of an an electric car can be beneficial for the environment. The significant differences in the share of renewable energies in the electricity mix should be quite visible.

Visualization 2: EVs Model Overview (Treemap / Tree Diagramm)

- Treemap
  - Filter between HEV/PEV
  - List selector for displayed data dimension
- Tree Diagram
  - Node selection -> expand tree branch
  - Hover over final node -> show Model statistics

This visualisation will give the reader the possibility to explore the model variety of Electric Vehicles. He will furthermore be able to compare the models by specific data dimension like fuel economy to obtain a more diverse overview and hopefully contribute to a decision finding.

Visualization 3 : HEV/PEV Sales statistics (treemap)

- Filter by Manufacturer
- Timeline range slider for annual sales statistics

This visualisation will give an overview about the sales of Electric Vehicles. It will provide the reader with a tool to discover developments and trends on the market, to ideally help the opinion finding process.

Visualization 4: Electric charger network

- Postal code entry box -> geographical limiter (circle) to range slider radius
  Alternatively : click on map generates postal code
- Range slider -> increase / decrease the radius of the geographical limiter
- Electric charger type distribution diagram -> linked to currently selected region (updates when limited)

3 Timeline

3.1 Milestones

Date	Milestone	Detail
28. March	Project Plan	Project proposal & detailed project plan
04. April	Final Project Plan	Finalizing of the project plan.
11. April	Project re-design	Re-Design according to inputs
18. April	Project prototype V1	First Prototype
25. April	Project prototype V2	Review
2. May	Final project V2	Project Demo / Screencast

3.2 To-Do List

Date	Task	Category	Responsible
13.04.2016	Electric Charger Types	Data Wrangling	Byron
13.04.2016	Reorganisation of Github	Infrastructure	Tim
16.04.2016	V1: Electric Charger Network	Visualisation	Enrico
16.04.2016	V1: Choropleth in D3 (CO2 Equivalence)	Visualisation	Byron
17.04.2016	V1: Linked Barchart in D3 (CO2 Equivalence)	Visualisation	Byron
16.04.2016	V1: Treemap in D3	Visualisation	Tim
17.04.2016	V1: Linked View Cars with (price and image)	Visualisation	Tim
18.04.2016	V1: Website Design	Presentation	Tim
18.04.2016	V1: Binary Decision Tree (w/o links)	Visualisation	Enrico
18.04.2016	Additional text	Content	Tim

4 Team Roles

4.1 Necessary Skills

- Project Management
- Design skills (Layout, Pictures etc.)
- Data Manipulation
- Web Scrapping
- API Connection
- Visualization
- HTML/CSS

4.2 Team Roles

Coordination, Coding & Data: Tim Hagmann

Sketching, Coding & Data: Enrico Mund

Coding & Data: Byron Bahan

5 Homepage

5.1 Storytelling Questions / Feature list

These question should be answered for the reader

The EV - What is it ?
How does it work ? (technology)
How does it work for the environment ? (renewable energies / sources of electricity)
Where is the electric car beneficial for the environment?
What can I buy ? (models)
What do other people buy ? (sales)
Where will I charge my new car / get service?

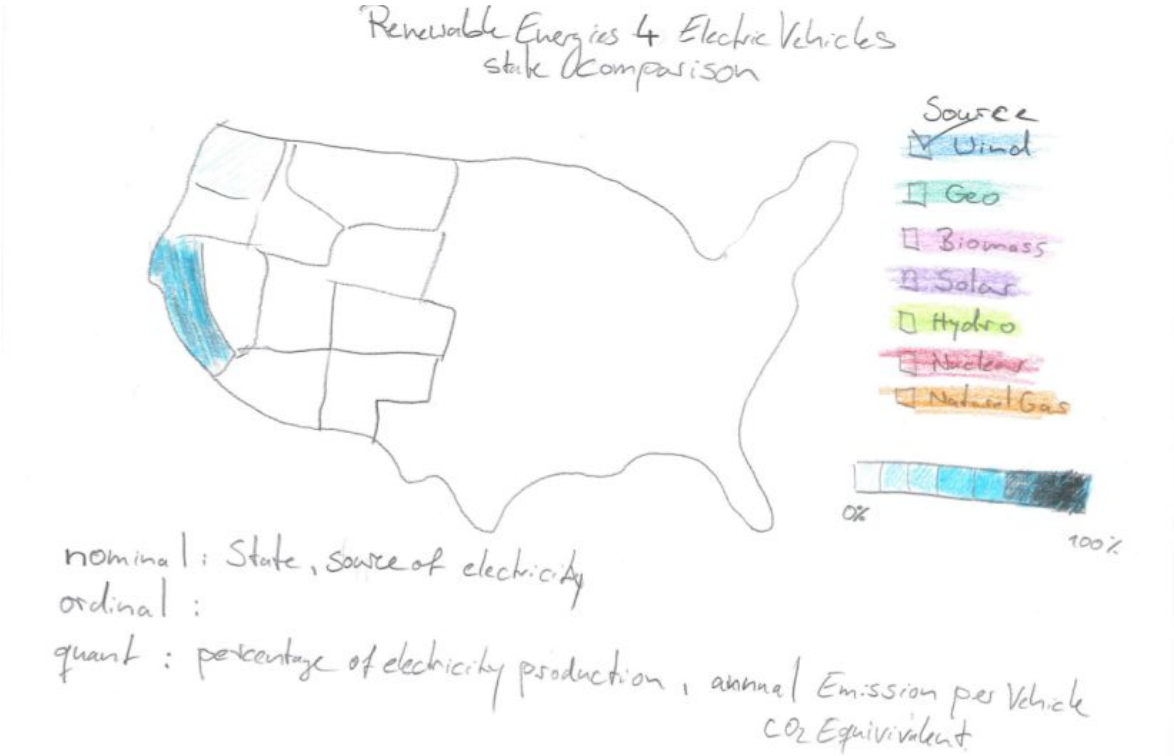
The final project include the following features:

- Implemented multiple coordinated linked views
- A novel visualization

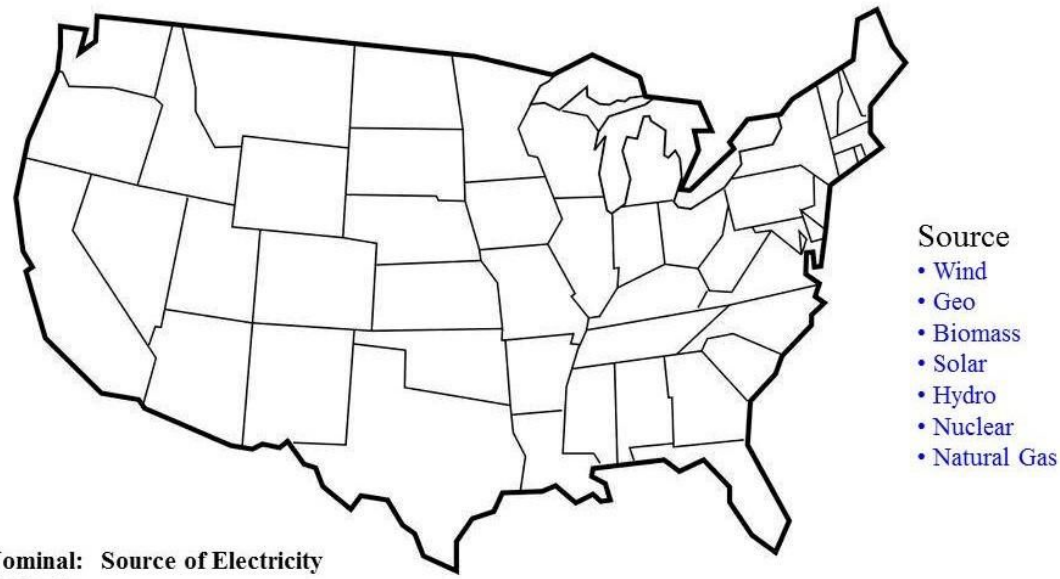
All visualizations have the goal to be:

- Effective
- Innovative
- Clear and aid the storytelling
- Include sensible and effective interactions
- Goal oriented

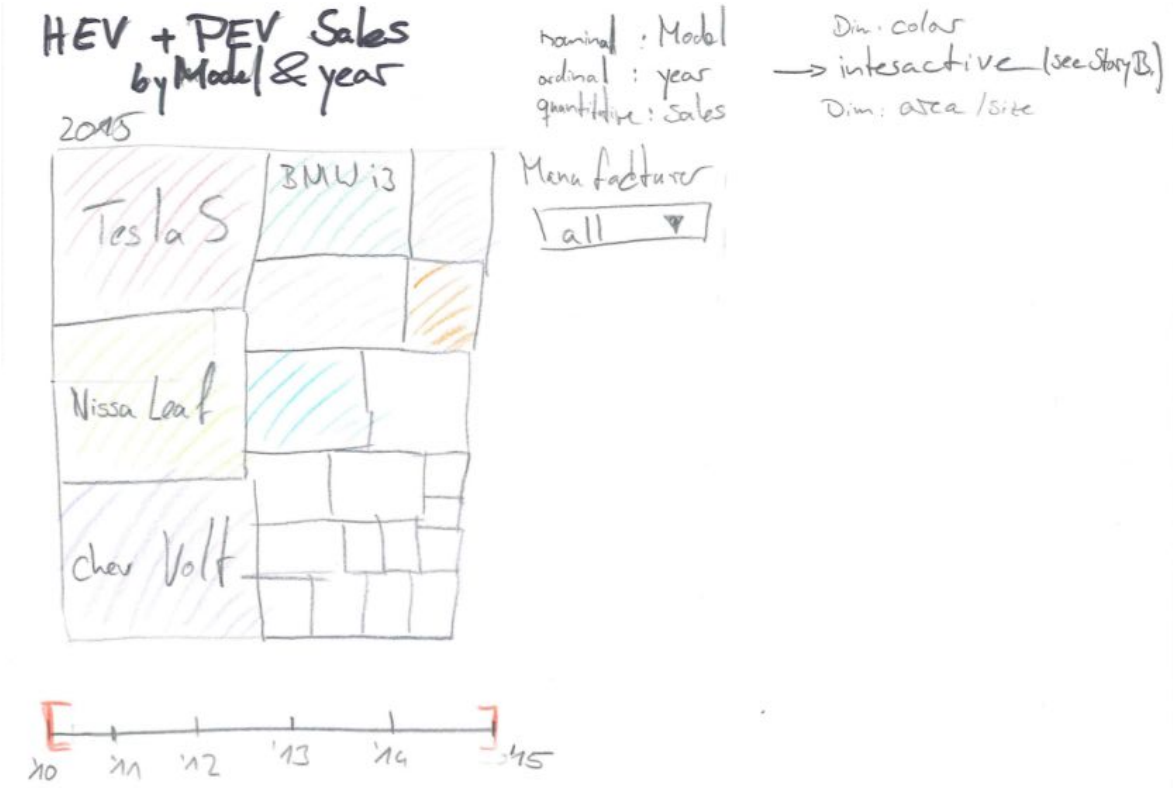
5.2 2nd iteration of Sketches

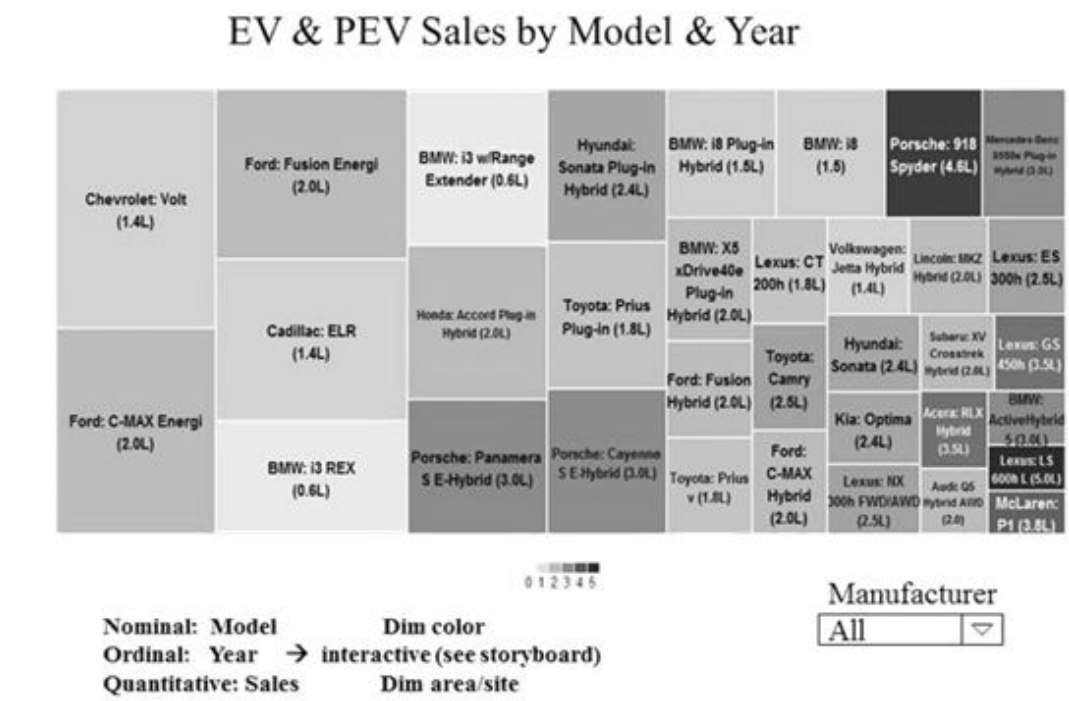


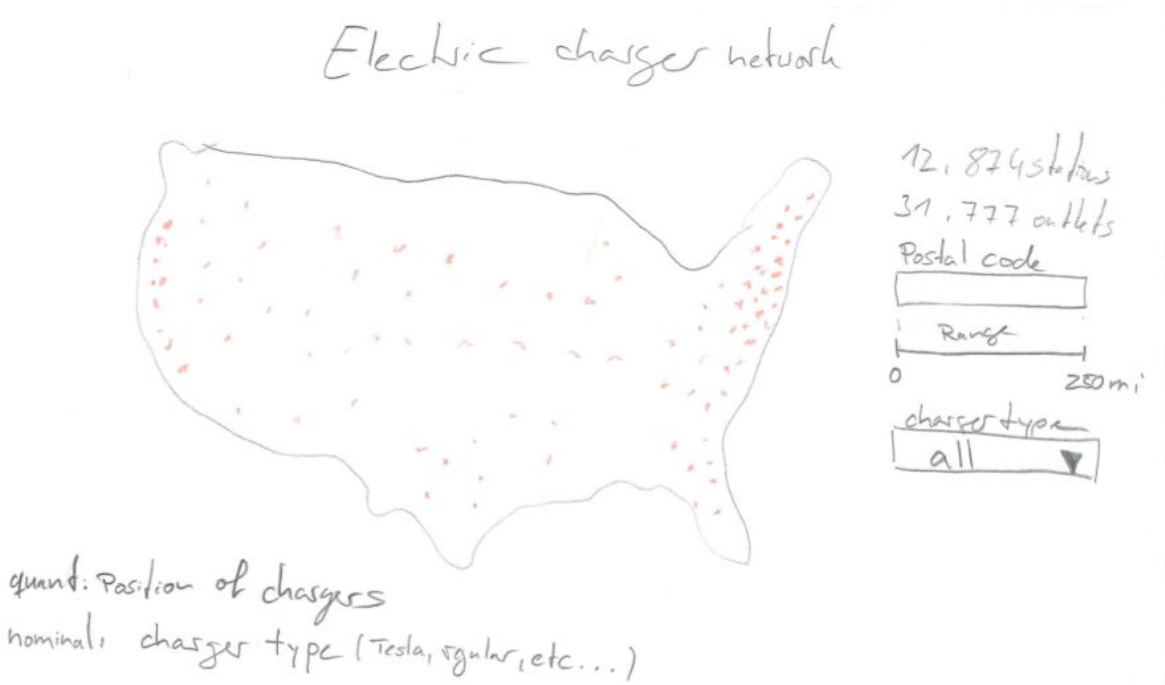
Renewable Energies for Electric Vehicles
State by State Comparison

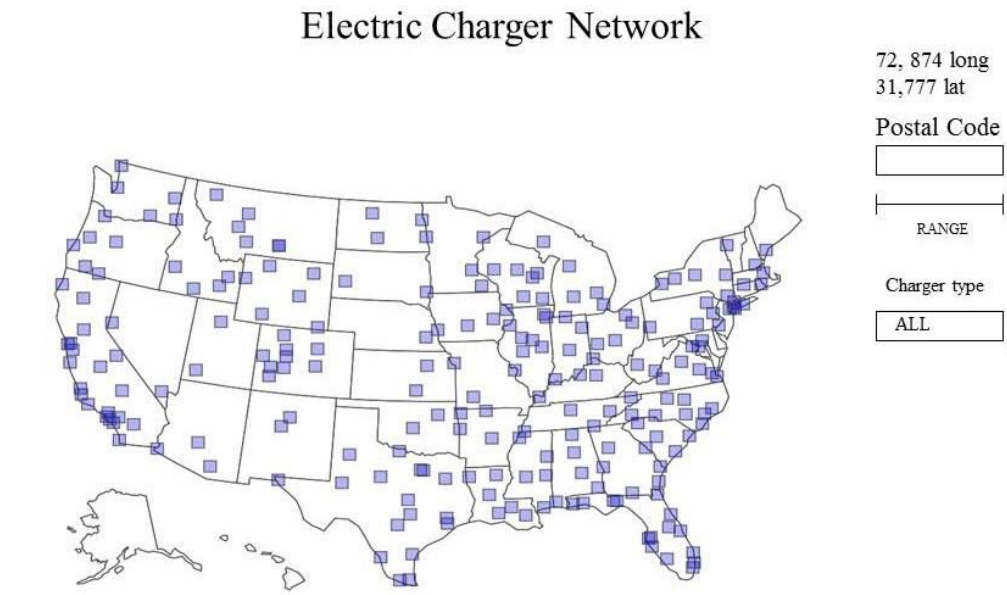


Nominal: Source of Electricity
Ordinal:
Quant: Percentage of Electricity Production, Annual Emission per Vehicle CO2 Equivalent

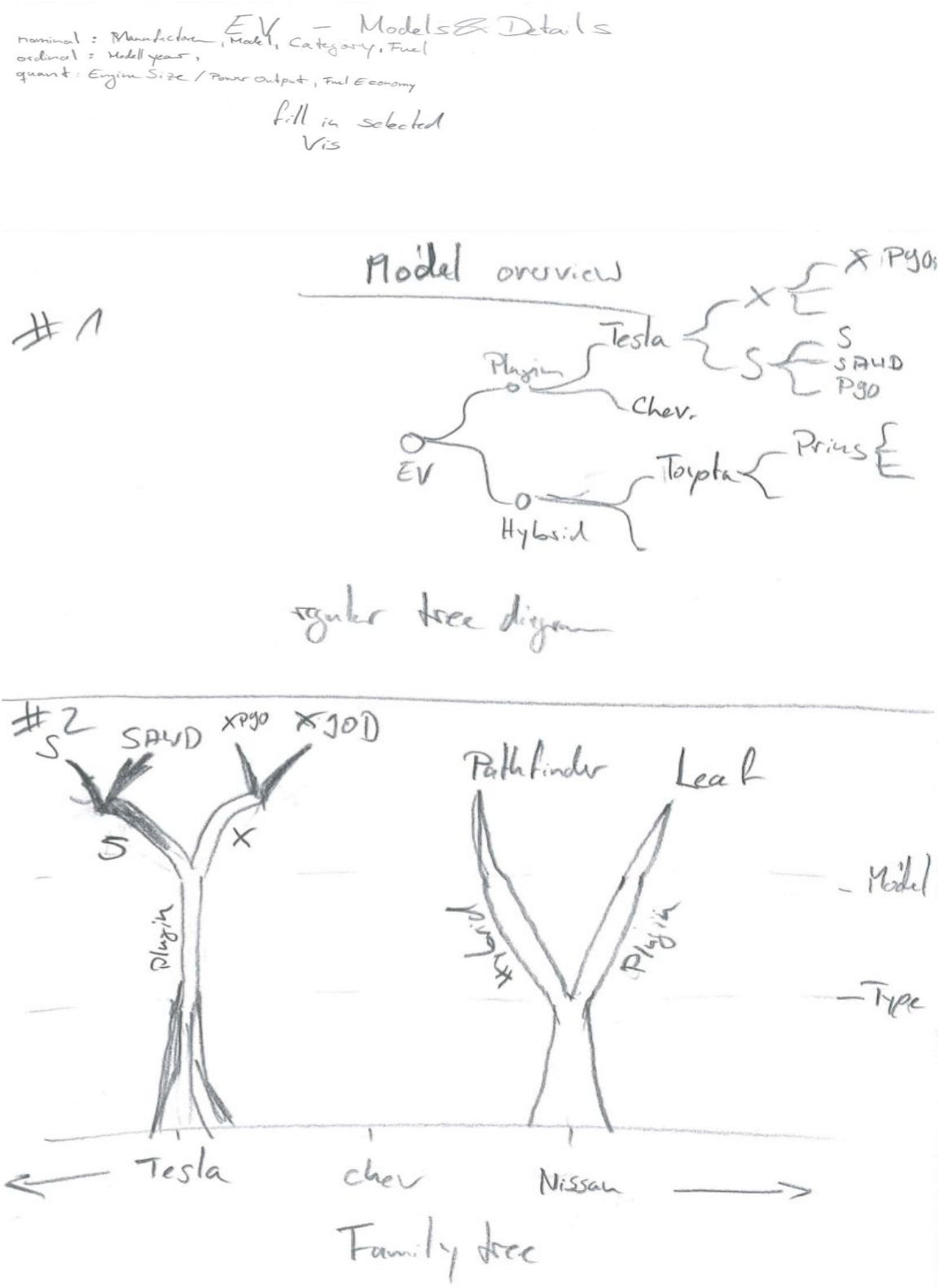

HEV + PEV Sales
by Model & year
nominal : Model
ordinal : year
quantitative : Sales
Dim: color
→ interactive (see Story B.)
Dim: area / site
2015
Tesla S
BMW i3
Nissa Leaf
Chev Volt
Manufacturer
all
'10
'11
'12
'13
'14
'15

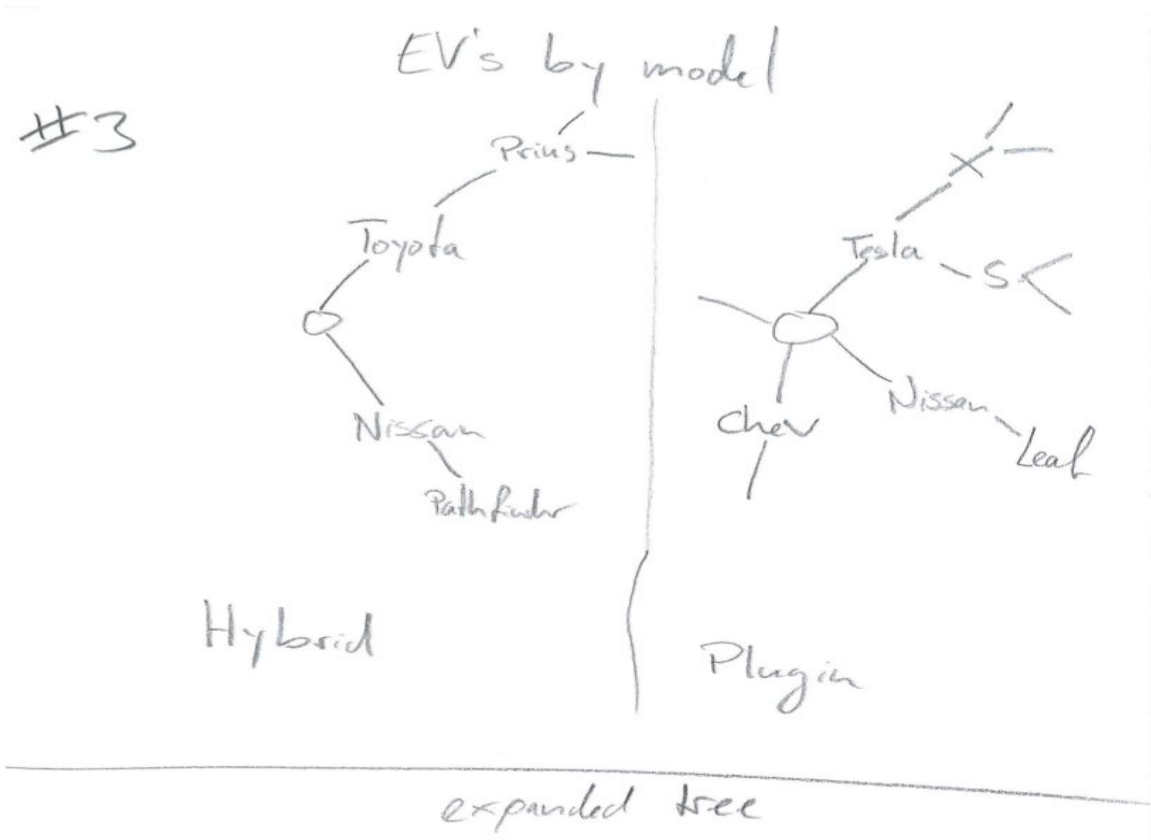

EV & PEV Sales by Model & Year
Chevrolet: Volt (1.4L)
Ford: Fusion Energi (2.0L)
BMW: i3 w/Range Extender (0.6L)
Hyundai: Sonata Plug-in Hybrid (2.4L)
BMW: i8 Plug-in Hybrid (1.5L)
BMW: i8 (1.5)
Porsche: 918 Spyder (4.6L)
Mercedes-Benz S550e Plug-in Hybrid (3.0L)
BMW: X5 xDrive40e Plug-in Hybrid (2.0L)
Lexus: CT 200h (1.8L)
Volkswagen: Jetta Hybrid (1.4L)
Lincoln: MKZ Hybrid (2.0L)
Lexus: ES 300h (2.5L)
Cadillac: ELR (1.4L)
Honda: Accord Plug-in Hybrid (2.0L)
Toyota: Prius Plug-in (1.8L)
Ford: Fusion Hybrid (2.0L)
Toyota: Camry (2.5L)
Hyundai: Sonata (2.4L)
Subaru: XV Crosstrek Hybrid (2.0L)
Lexus: GS 450h (3.5L)
Ford: C-MAX Energi (2.0L)
BMW: i3 REX (0.6L)
Porsche: Panamera S E-Hybrid (3.0L)
Porsche: Cayenne S E-Hybrid (3.0L)
Toyota: Prius v (1.8L)
Ford: C-MAX Hybrid (2.0L)
Kia: Optima (2.4L)
Acura: RLX Hybrid (3.5L)
BMW: ActiveHybrid 5 (3.0L)
Lexus: LS 600h L (5.0L)
Lexus: NX 300h FWD/AWD (2.5L)
Audi: Q5 Hybrid AWD (2.0)
McLaren: P1 (3.8L)
0 1 2 3 4 5
Manufacturer
All
Nominal: Model
Ordinal: Year
Quantitative: Sales
Dim color
→ interactive (see storyboard)
Dim area/site

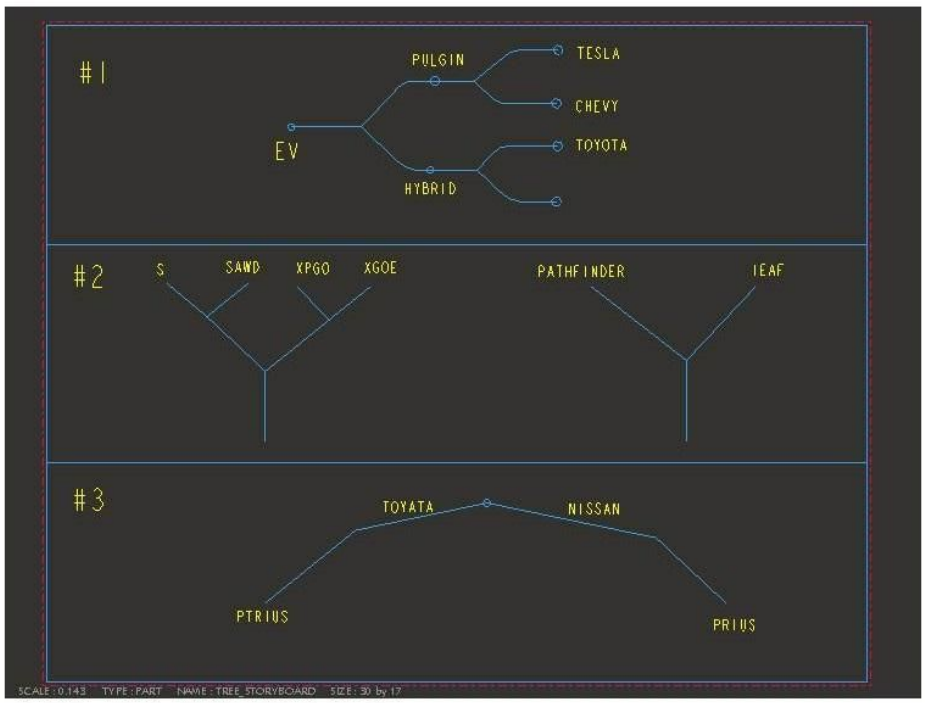




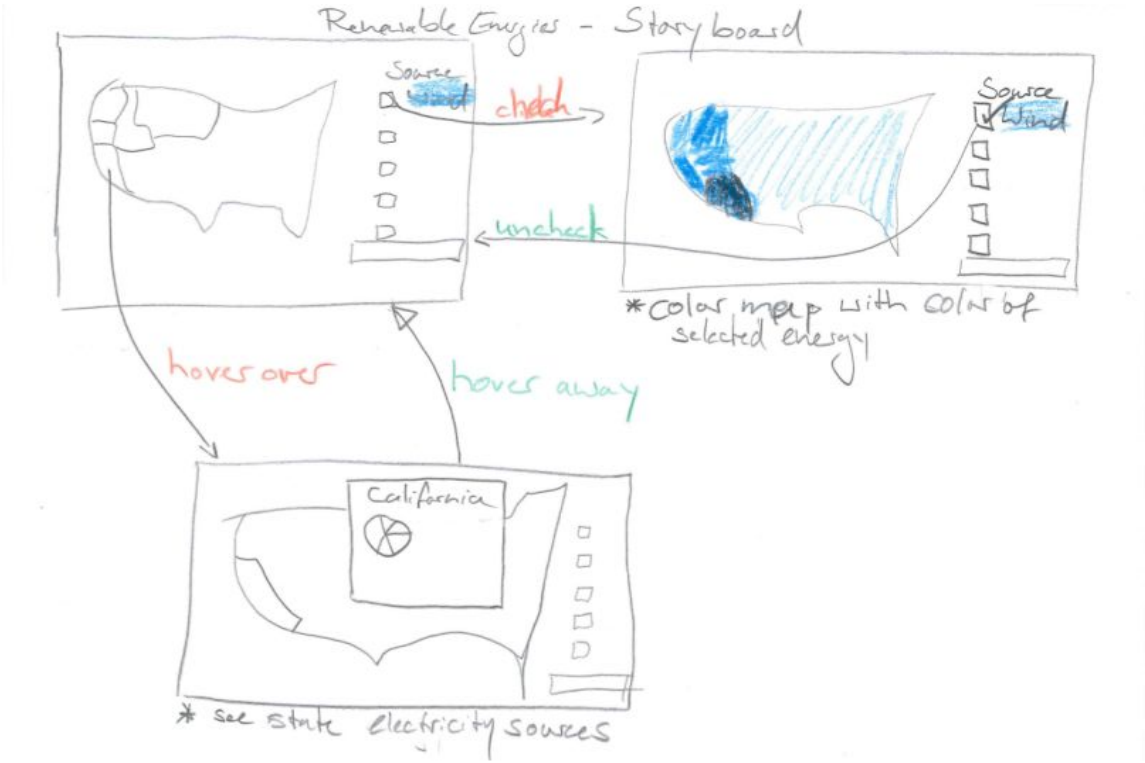
Quant: Positions of Chargers
Nominal: Charger Type per MFG

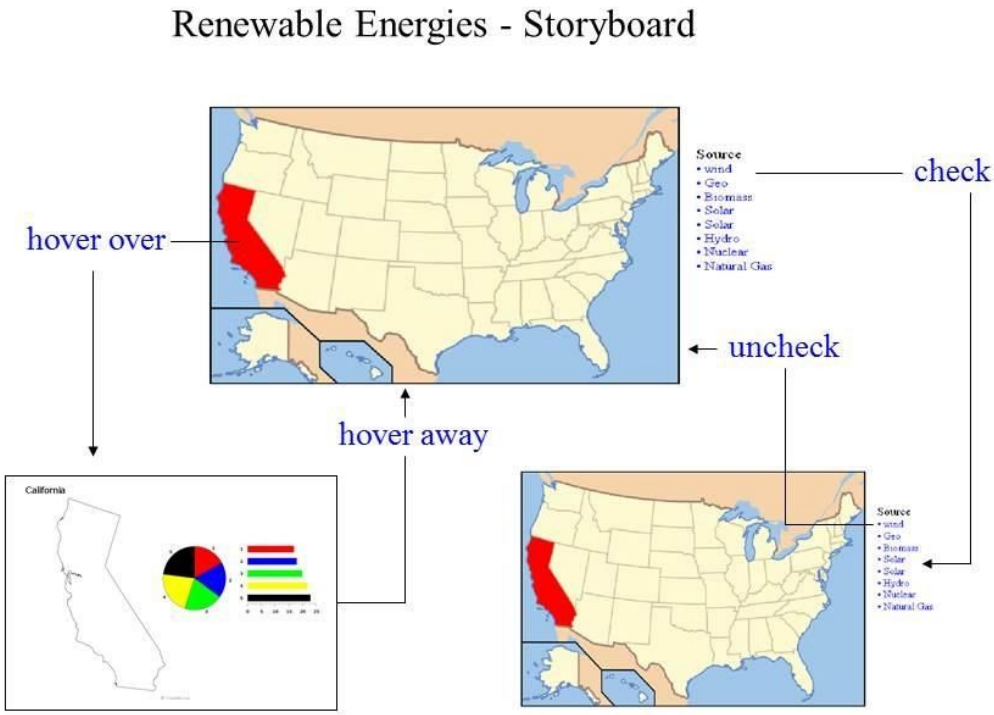

EV - Models & Details
nominal : Manufacturer, Model, Category, Fuel
ordinal : Model year,
quant : Engine Size / Power output, Fuel Economy
fill in selected
Vis
Model overview
#1
Tesla
Plugin
Chev.
EV
Hybrid
Toyota
Prius
S
SAUD
PGO
regular tree diagram
#2
S
SAUD
XPGO
XJOD
Pathfinder
Leaf
Model
Type
Tesla
chev
Nissan
Family tree

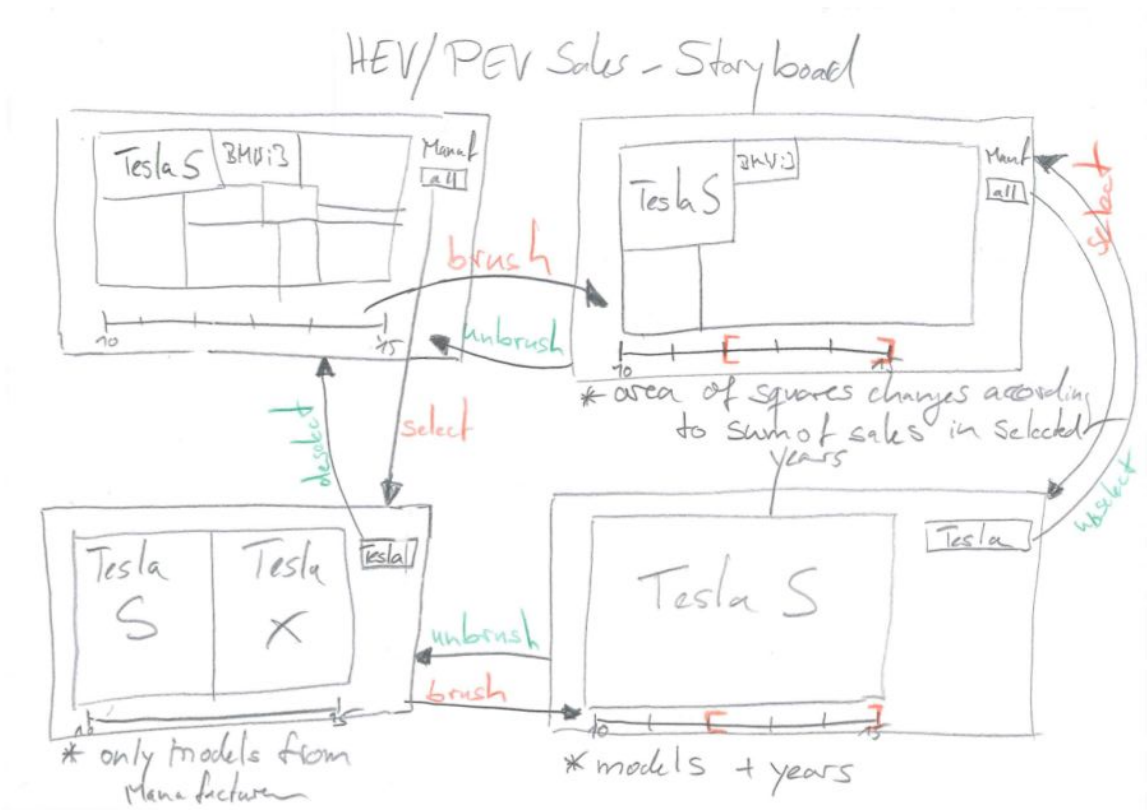

#3
EV's by model
Prius
Toyota
Nissan
Pathfinder
Hybrid
Tesla
S
Chev
Nissan
Leaf
Plugin
expanded tree


#1
EV
PULGIN
HYBRID
TESLA
CHEVY
TOYOTA
#2
S
SAWD
XPGO
XGOE
PATHFINDER
IEAF
#3
TOYATA
NISSAN
PTRIUS
PRIUS

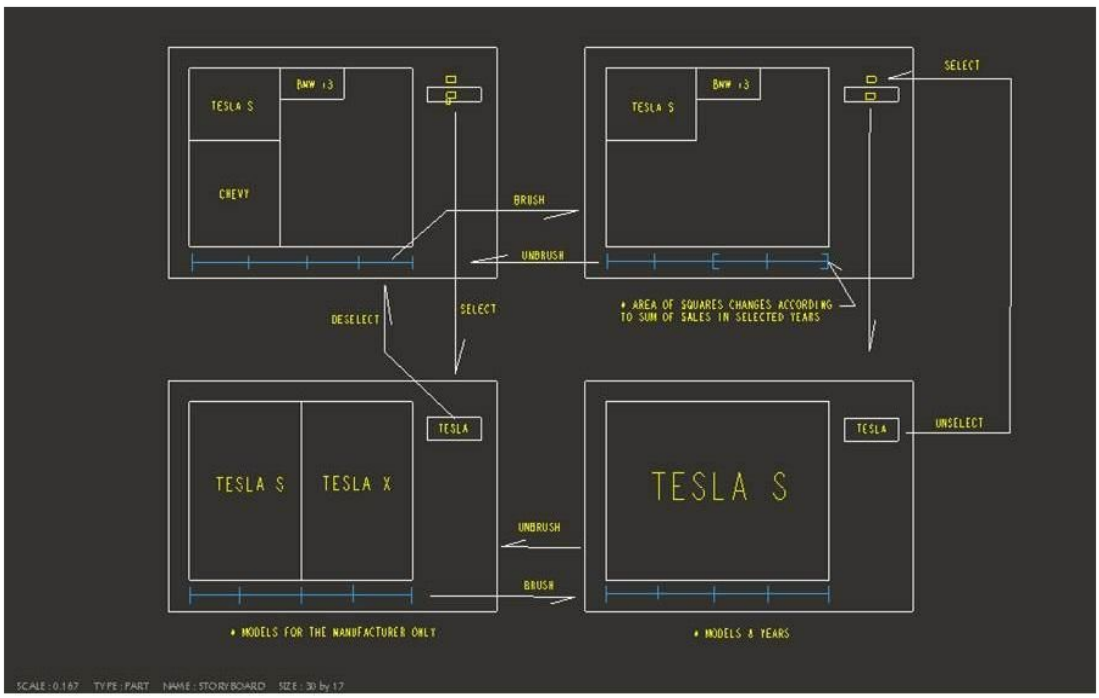
5.3 Storyboard



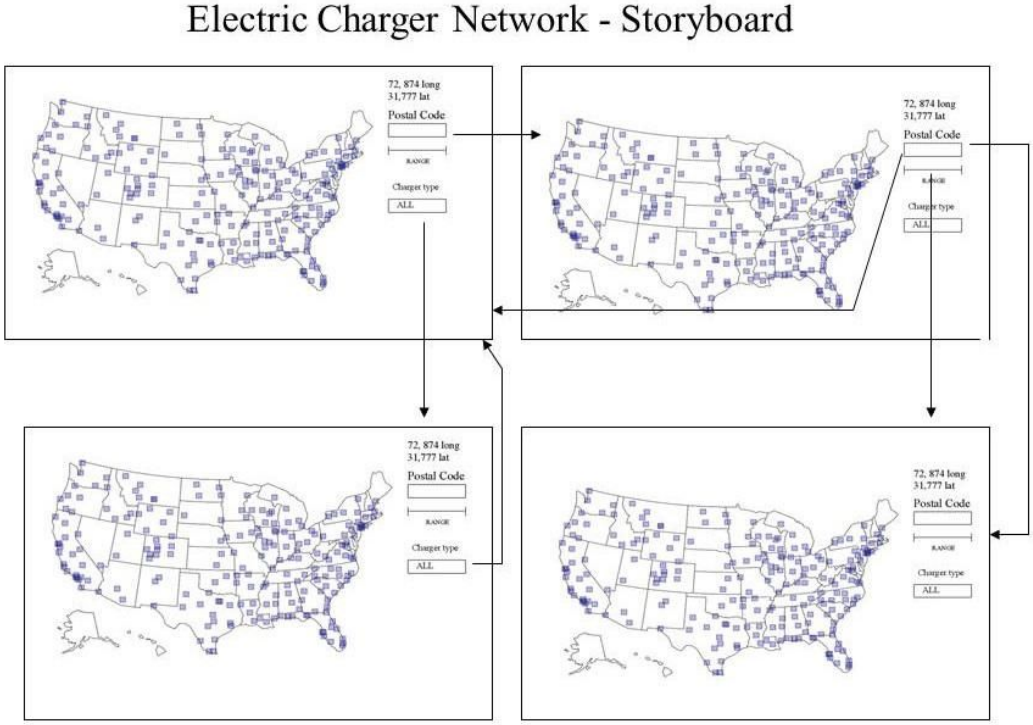




HEV/PEV SALES - Storyboard



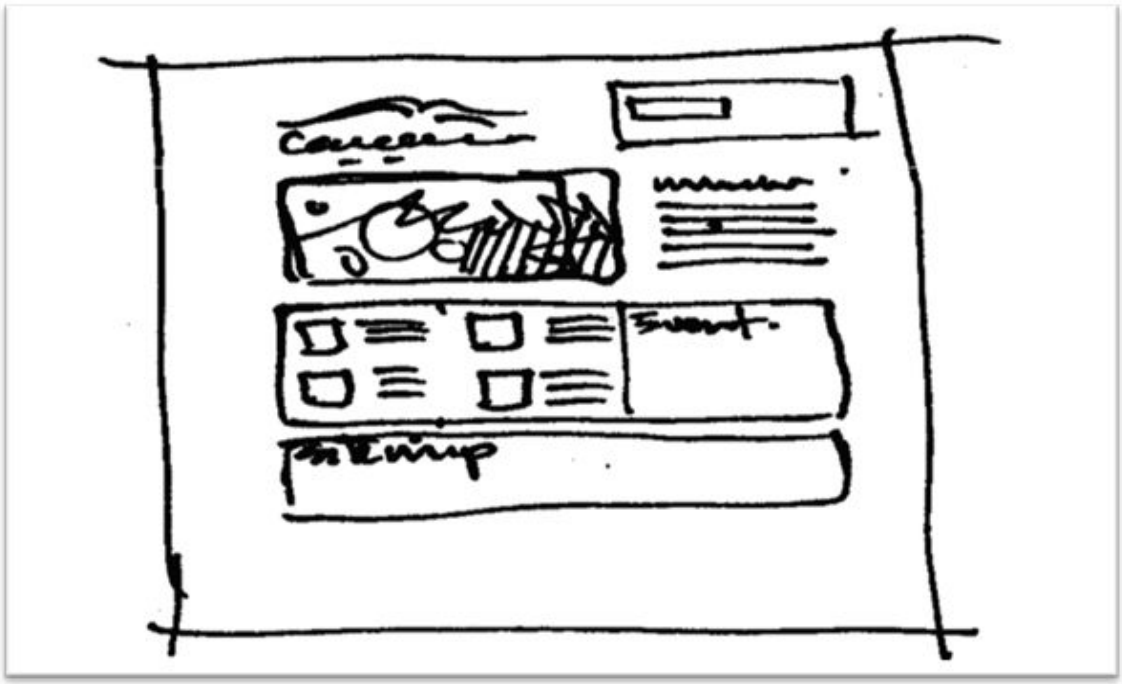

Electric charger network – Storyboard
12,874
Pcode
Range
type
all
Enter
remove
Select
Release
1000
3400
250
all
* circle around postal code location
brush
unbrush
5.000
7,450
Pcode
Range
type
Tesla
500
1250
100
all
* smaller circle based on range


Electric Charger Network - Storyboard
72, 874 long
31,777 lat
Postal Code
RANGE
Charger type
ALL

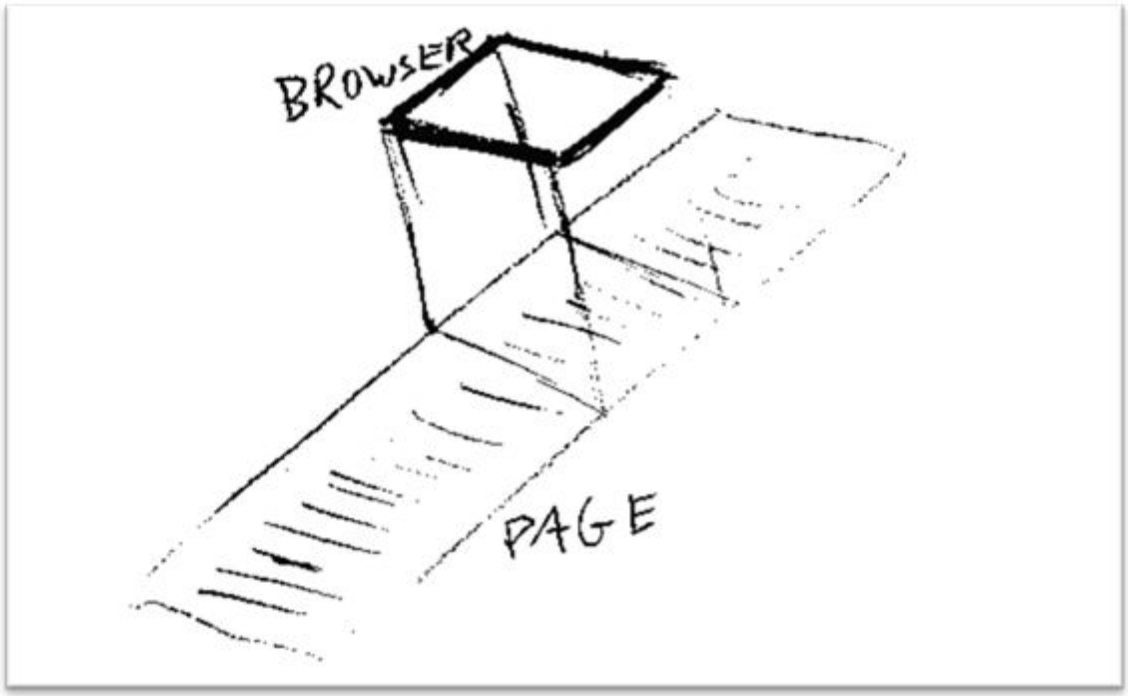
5.4 Webpage Layout

We are using Bootstrap default layouts and are following a vertical scrolling approach. That means that all the information is on one page and the user is scrolling to the page to get to the information.

5.4.1 Bootstrap layout (Main Page)



5.4.2 Vertical scrolling (Page Logic)



6 Redesign

In order to allow for a first user feedback a simple prototype was created (http://greenore.github.io/EV-Website). Then a heuristic evaluation was conducted that was based on three task. (The whole process was filmed and can be accessed through the following link: https://goo.gl/2uYkIC)

6.1 Heuristic Evaluation

Task 1: Storyline

In the first task the user had to "guess" what the storyline of the project is. The goals was to find out if at a short the webpage already shows the purpose of the project or if the storyline needs some refinement.

Result

- Needs a better title - "Team Tesla" suggests that the project is about Teslas.
    - Nielson Category: Help and Documentation
    - Severity Scale: 2

  → TeamTesla is going to move to the impressum and a shorter clearer title is being tested.

- Dense introduction - the introduction under "Motivation and Abstract" is too long.
    - Nielson Category: Help and Documentation
    - Severity Scale: 3

  → Probably bullet points are used instead

- Background color - Black text on map legend, for example, is not visible.
    - Nielsen Category: Aesthetic and Minimalist Design
    - Severity Scale: 2

  → Either a lighter background or some different colors in the visualizations are being tested.

Task 2: Fuel Economy Chart

The second task was to find a car that the user likes and to get the price information. The goals was to see if the treemap is intuitive.

Result

- The chart is confusing.

  → In order compare electric vehicles and hybrids, it would be helpful to see data regarding these two types of vehicles side-by-side.

  → Clarify what color represents (here, it is representing a quantitative metric (horse power), but color is generally weak for illustrating quantitative metrics.)

  → Clarify what the box-size represents

Task 3: Electric Charger Network

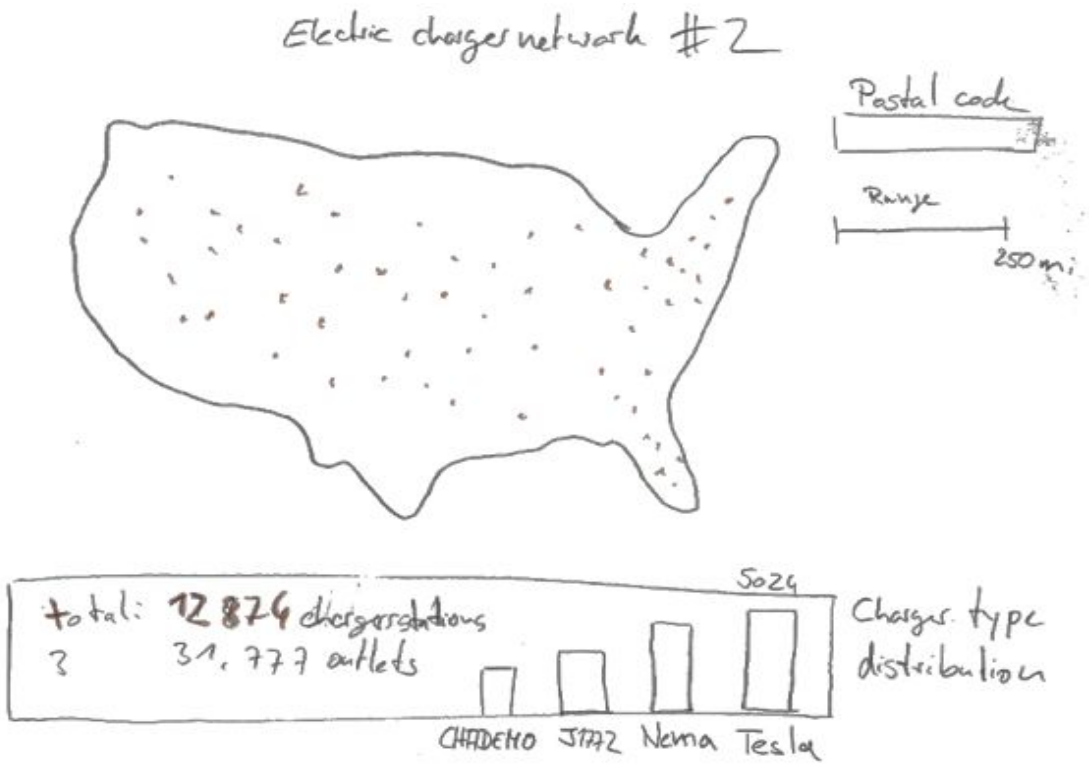
In the third task was to analyse the electric charger network and to get general input on what users expect from this graph.

Result

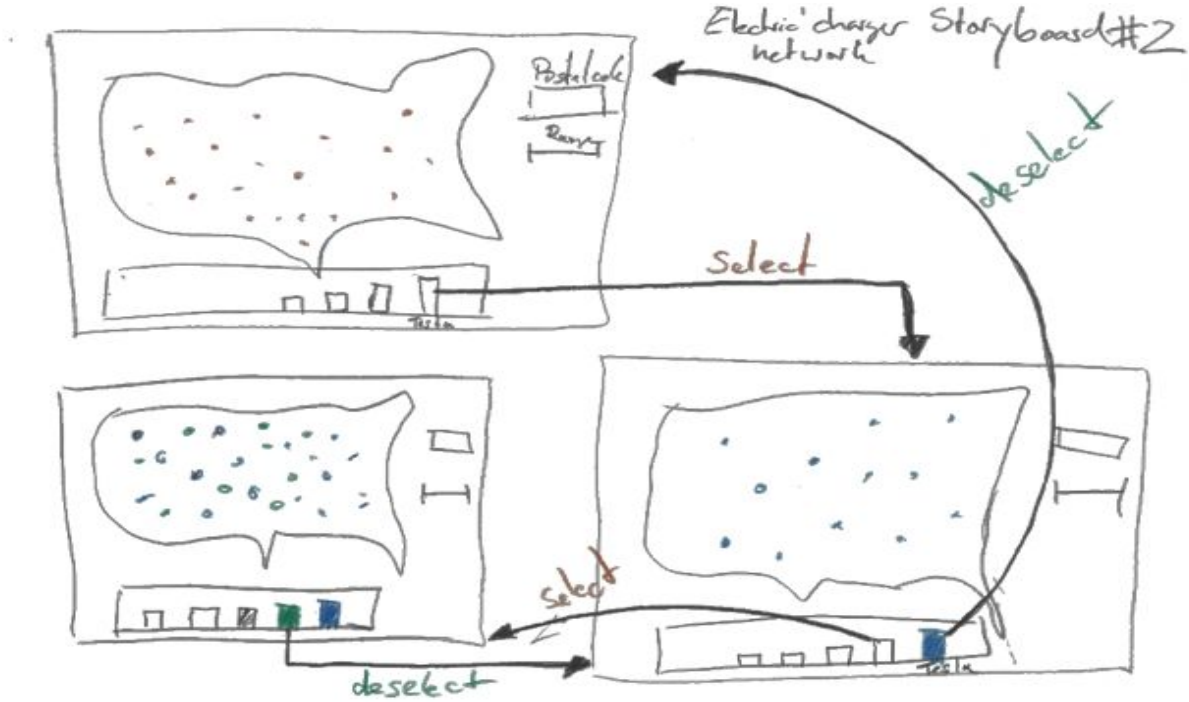
→ Make an option for user to add start and destination locations, and routes, with charging stations along that route.

→ Option to pick different car types and integrate that variable into the itinerary.

Sketch - Electric charger network #2 redesign



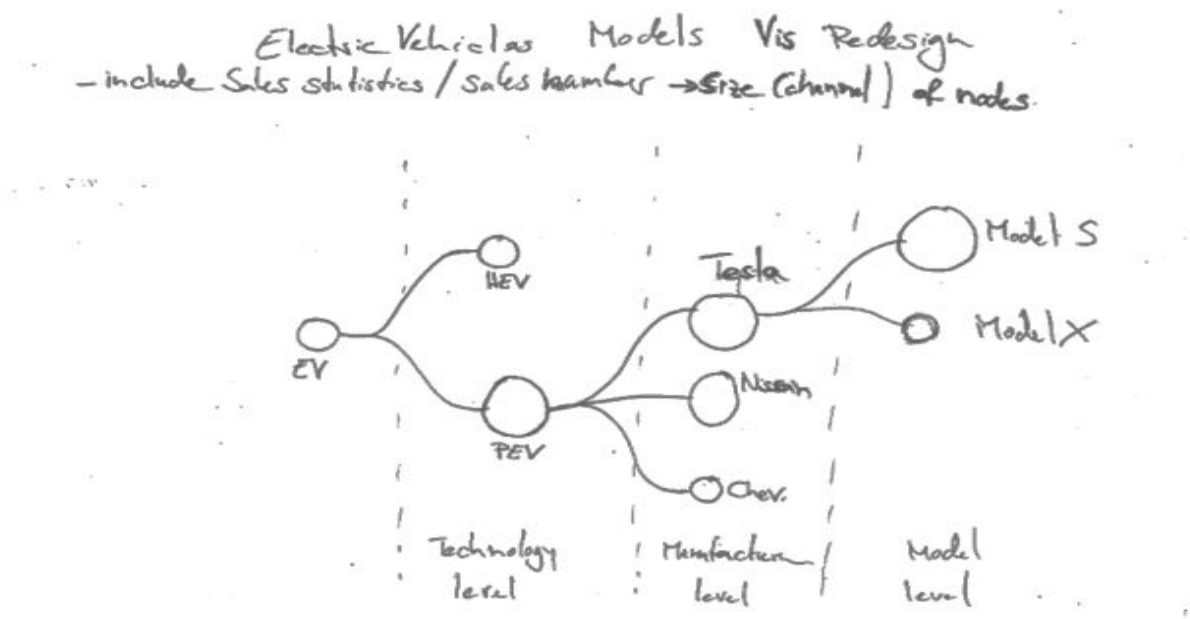
Sketch - Electric charger network #2 - Storyboard



6.2 Innovation

In order to facilitate the model choice for the user, we took a close look at options to visualize this. Our solution is a redesigned binary decision tree. The idea is to add additional channels to the tree. More precisely, we are varying the size of the decisions nodes according to the sales numbers of the cars. This gives the user a fast overview of the popularity of a specific car/car manufactures.

Tree diagram with quantitative dimension



7 Data

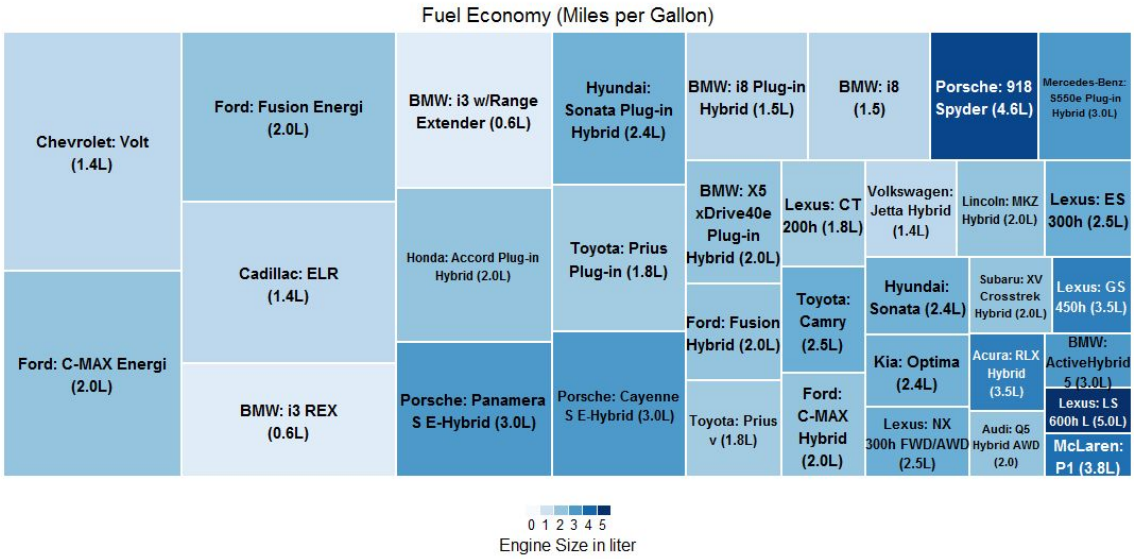
The primary data source will be http://www.afdc.energy.gov/data/ from the US Department of Energy. We have acquired sufficient data to generate at least 2 different visualizations without much data cleaning.

We will rely on an API provided by http://developer.nrel.gov/docs/transportation/alt-fuel-stations-v1/ to implement a third visualization. The API is very well documented and shouldn't require a great amount of work to obtain useful data.

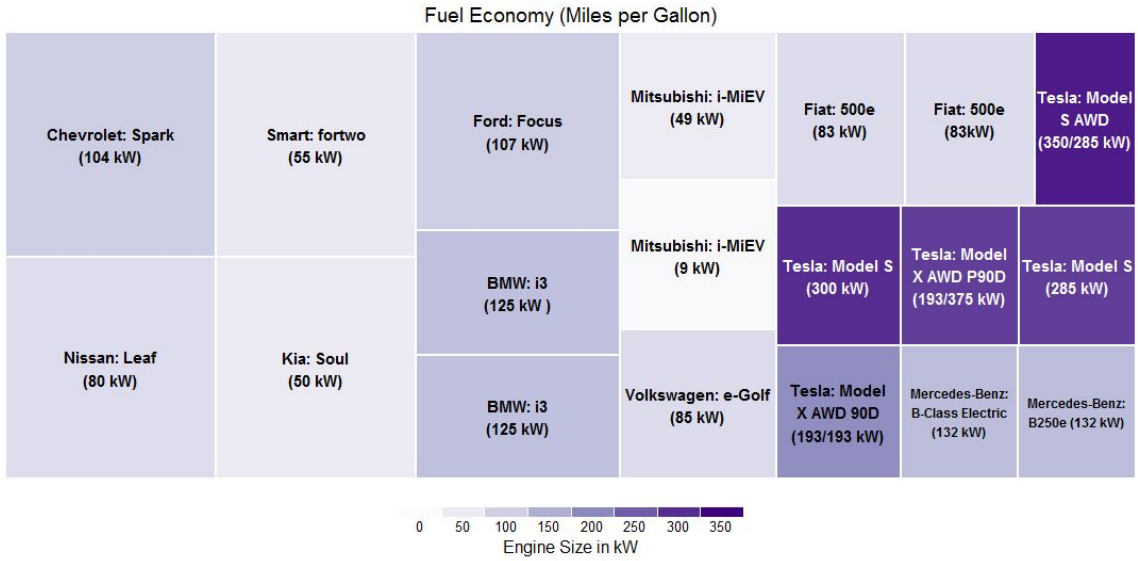
A fourth visualization will require some web-scraping from http://www.afdc.energy.gov/vehicles/electric_emissions.php. This data source will most likely require the largest amount of data munging, because it is not documented and will rely on reverse engineering. However we are confident that we can solve this within a reasonable time limit.

We also did some preliminary data analysis and data munging on one of the datasets. That is, the "Electric Vehicles by Model" set. The reproducible R-Code as well as the data is on a private Github repository: https://github.com/greenore/cs171-project

Fuel Economy for Hybrids



Fuel Economy for Electric Cars



APPENDIX

A1 Poster

Tree & Network Charts

(by Byron Bahan, Enrico Mund and Tim Hagmann)

Treemap

Source: http://goo.gl/2Es7mY

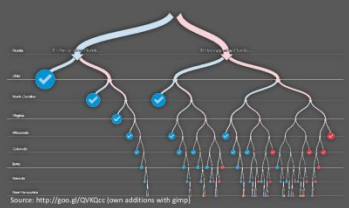
Definition
A treemap is a visualization method for hierarchical data using nested geometric elements. These elements are usually rectangles but other shapes can be implemented.

Idea and implementation
Tree structured data are drawn using a tiling algorithm. Each data branch is given a rectangle (tile) and can then be tiled into sub-branches using smaller tiles. Size and color dimensions of these tiles usually correspond to specific data from the data set. Selection of the color dimension is of significance since patterns can be recognized easier when color scales are well picked. Rectangular treemaps make efficient use of space and can display large quantities of items.

Tiling algorithms
Ordering of the elements and maintaining their aspect ratio is difficult to achieve as these properties have an inverse relationship.

Algorithm	Order	Aspect Ratio
Binary tree	partially ordered	high
Mixed Treemaps	ordered	lowest
Ordered	partially ordered	medium
Squarified	unordered	lowest

Decision Tree



Source: http://goo.gl/i2VKQcc (own additions with gimp)

Definition
A decision tree is a flowchart-like structure in which each node represents a "test/split" on an attribute. Each branch represents the outcome of this test/split and each leaf represents a class label.

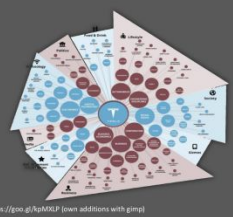
Idea and implementation
A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has all the same value of the target variable, or when splitting no longer adds value to the predictions.

There are two main types of decision trees:
- Classification trees → predicted outcome is a class
- Regression trees → predicted outcome is a number

Splitting algorithms
There are many specific decision-tree algorithms. Notable ones include:

Algorithm	Splits	Metric
CART	Binary	Gini impurity
CHAID	Binary	Chi-Squared
C5.0	Multiple	Information gain
MARS	Binary	GCV Error

Node-Link Chart



Source: https://goo.gl/kpMXLP (own additions with gimp)

Definition
Lorem ipsum

Idea and implementation
Lorem ipsum

Algorithms
Lorem ipsum

Lorem ipsum	Dolor sit	Amet
...	...	...
...	...	...
...	...	...
...	...	...

A2 Vis Exploration Questionnaire

CS 171

Vis Exploration Poster Session

Please submit this questionnaire - one sheet for each team member - and a photo of your poster (one per team) with your homework

The timetable gives you an overview which expert group has to present and which expert group will go around and explore visualization techniques. Please stand next to your poster when your expert group is presenting.

		Presenting	Exploring
2:45 - 3:00	A — Geographical Data / Maps	A	B C D
3:00 - 3:15	B — Trees & Networks	B	A C D
3:15 - 3:30	C — Visualization for Text	C	A B D
3:30 - 3:45	D — Vis for High Dimensional Data	D	A B C
3:45 - 4:00	Vote for your best poster		

Best Poster Award: Please put your sticky dots on the posters that you like most. You can put multiple dots on the same poster if you wish to do so.
One Minute Paper: Please do not forget to submit the one minute papers! You have until tomorrow.

Name: Tim Hagmann

Group name: Team Tesla

List **three** visualization techniques you found most commonly on the posters about Geographical Data

- Choropleth
- Heat Map
- Sunburst plot

CS 171

List **three** visualization techniques you found most commonly on the posters about Trees and Networks

- Node-Link
- Treemaps
- Binary Tree

List **three** visualization techniques you found most commonly on the posters about Text Visualization

- Word Cloud
- Word Network
- Word Tree

List **three** visualization techniques you found most commonly on the posters about High Dimensional Data

- Parallel Coordinates
- Star Maps
- Horizon Graph

Next lecture: Evaluation & Innovation with D3

CS 171

Vis Exploration Poster Session

Please submit this questionnaire - one sheet for each team member - and a photo of your poster (one per team) with your homework

The timetable gives you an overview which expert group has to present and which expert group will go around and explore visualization techniques. Please stand next to your poster when your expert group is presenting.

		Presenting	Exploring
2:45 - 3:00	A — Geographical Data / Maps	A	B C D
3:00 - 3:15	B — Trees & Networks	B	A C D
3:15 - 3:30	C — Visualization for Text	C	A B D
3:30 - 3:45	D — Vis for High Dimensional Data	D	A B C
3:45 - 4:00	Vote for your best poster		

Best Poster Award: Please put your sticky dots on the posters that you like most. You can put multiple dots on the same poster if you wish to do so.
One Minute Paper: Please do not forget to submit the one minute papers! You have until tomorrow.

Name: Byron Bahan

Group name: Team Tesla

List **three** visualization techniques you found most commonly on the posters about Geographical Data

Pie Chart
Scattered Plots
Data Layers

CS 171

List **three** visualization techniques you found most commonly on the posters about Trees and Networks

MESH
STAR
TREE

List **three** visualization techniques you found most commonly on the posters about Text Visualization

Bamboo
ManyEyes
WordSmith

List **three** visualization techniques you found most commonly on the posters about High Dimensional Data

Biological Data
Random Field Data
Collaborative Filtering Data

Next lecture: Evaluation & Innovation with D3

CS 171

Vis Exploration Poster Session

Please submit this questionnaire - one sheet for each team member - and a photo of your poster (one per team) with your homework

The timetable gives you an overview which expert group has to present and which expert group will go around and explore visualization techniques. Please stand next to your poster when your expert group is presenting.

		Presenting	Exploring
2:45 - 3:00	A — Geographical Data / Maps	A	B C D
3:00 - 3:15	B — Trees & Networks	B	A C D
3:15 - 3:30	C — Visualization for Text	C	A B D
3:30 - 3:45	D — Vis for High Dimensional Data	D	A B C
3:45 - 4:00	Vote for your best poster		

Best Poster Award: Please put your sticky dots on the posters that you like most. You can put multiple dots on the same poster if you wish to do so.
One Minute Paper: Please do not forget to submit the one minute papers! You have until tomorrow.

Name: Enrico Mund

Group name: Team Tesla

List **three** visualization techniques you found most commonly on the posters about Geographical Data

- Heat Map
- Flow Map
- Choropleth

CS 171

List **three** visualization techniques you found most commonly on the posters about Trees and Networks

- Treemap
- Sunburst Plot
- Sankey Diagramm

List **three** visualization techniques you found most commonly on the posters about Text Visualization

- Word Network
- Word Tree
- Word Cloud

List **three** visualization techniques you found most commonly on the posters about High Dimensional Data

- Three Dimensional
- Parallel Coordinates
- Horizon Graph

Next lecture: Evaluation & Innovation with D3