Benchmarking Natural Language to Visualization Models

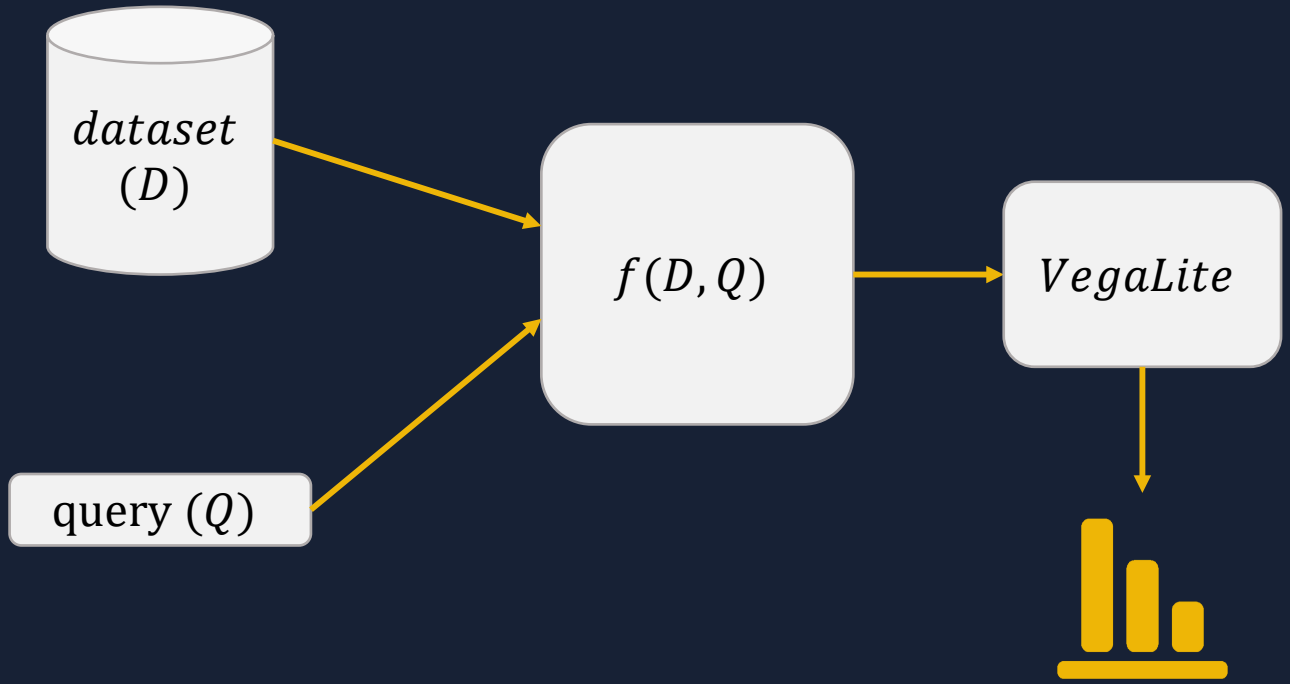
6.S079 Final Project

Enrique Casillas

Motivation

- Data visualization is important
- Turning a thought (natural language) into a visualization is the ultimate goal of visualization platforms

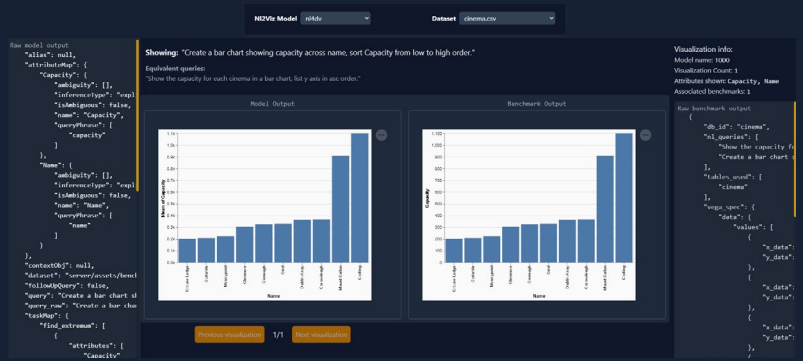
NL2Viz Models



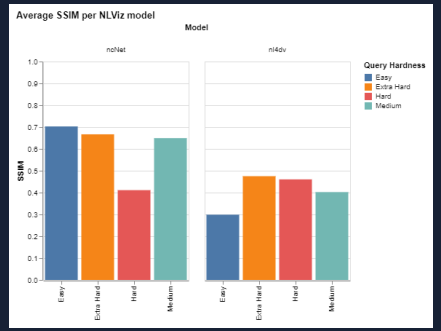
Project Overview

- Primary goal — benchmark NL2Viz models by using common metrics

Web-based tool – `nl2viz`



Evaluation of two models, `ncNet` and `nl4dv` using the `nvBench` NL2Viz benchmark



Web-based tool – `nl2viz`

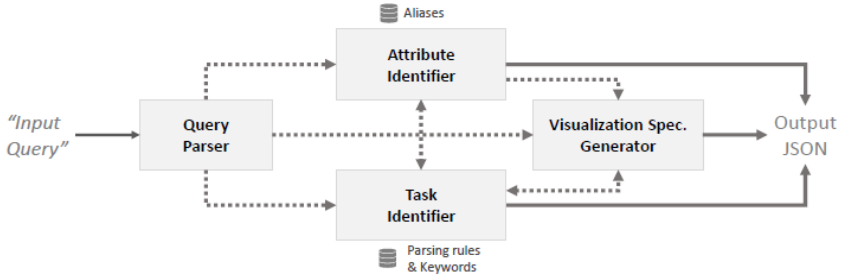
`nl2viz` Goals

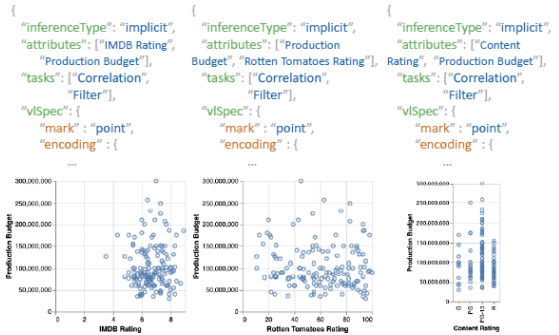
- Visually **compare** a model's output to a benchmark in a couple of clicks
- Get a feel for what the benchmark looks like and **expects**
- Help make your own **conclusions** as to which model is better
- Provide ***any*** query and receive the model's output

Demo

Model Evaluations

nl4dv

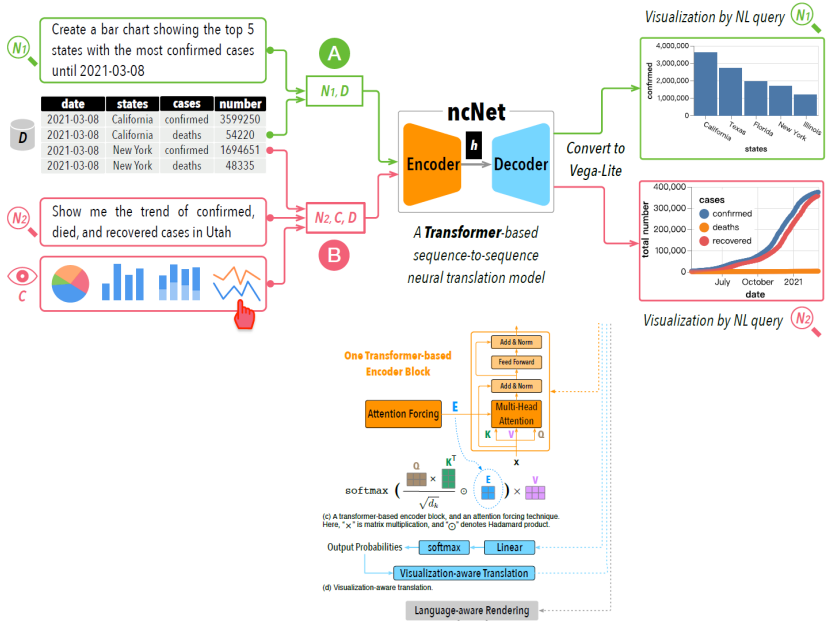




visList

Arpit Narechania, Arjun Srinivasan, and John Stasko. 2020. NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* (2020). https://doi.org/10.1109/TVCG.2020.3030378

ncNet



Yuyu Luo, Nan Tang, Guoliang Li, Jiawei Tang, Chengliang Chai, and Xuedi Qin. 2021. Natural Language to Visualization by Neural Machine Translation. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. https://doi.org/10.1109/TVCG.2021.3114848

Existing benchmark — **nvBench**


```
"vis_query": {
    "vis_part": "Visualize SCATTER",
    "data_part": {
        "sql_part": "SELECT FacID , count(*) FROM Faculty AS T1
JOIN Student AS T2 ON T1.FacID = T2.advisor GROUP BY T1.FacID",
        "binning": ""
    },
    "VQL": "Visualize SCATTER SELECT FacID , count(*) FROM Faculty
AS T1 JOIN Student AS T2 ON T1.FacID = T2.advisor GROUP BY
T1.FacID"
},
"chart": "Scatter",
"hardness": "Medium",
"db_id": "activity_1",
"vis_obj": {. . . },
"nl_queries": [. . .],
```


Yuyu Luo, Nan Tang, Guoliang Li, Chengliang Chai, Wenbo Li, and Xuedi Qin. 2021. Synthesizing Natural Language to Visualization (NL2VIS) Benchmarks from NL2SQL Benchmarks. (2021).

How do we compare visualizations?

Common Methods

Crowdsource

Ask Experts

Don’t evaluate

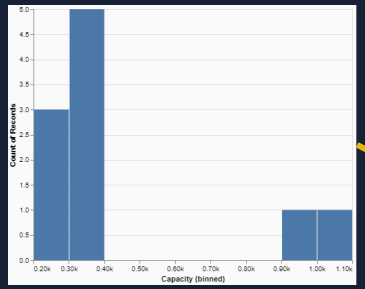
Alternative Methods

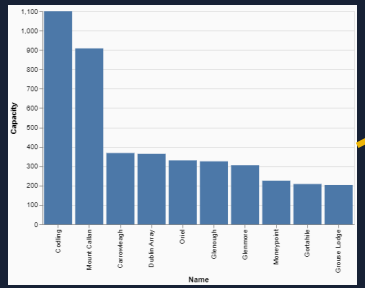
Compare specs

Image similarity

Structural Similarity Index Measure (SSIM)

nl4dv





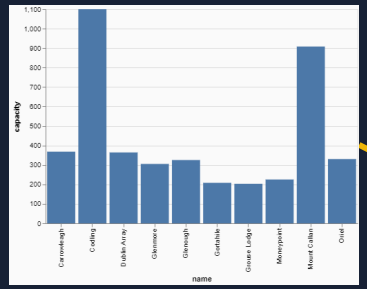
benchmark

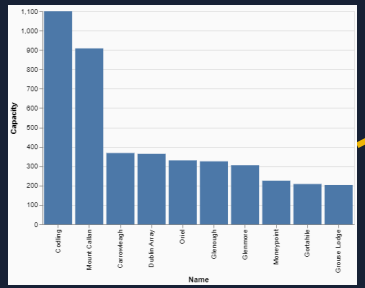
$SSIM(V_1, V_2)$

0.7164527617582078

Structural Similarity Index Measure (SSIM)

ncNet



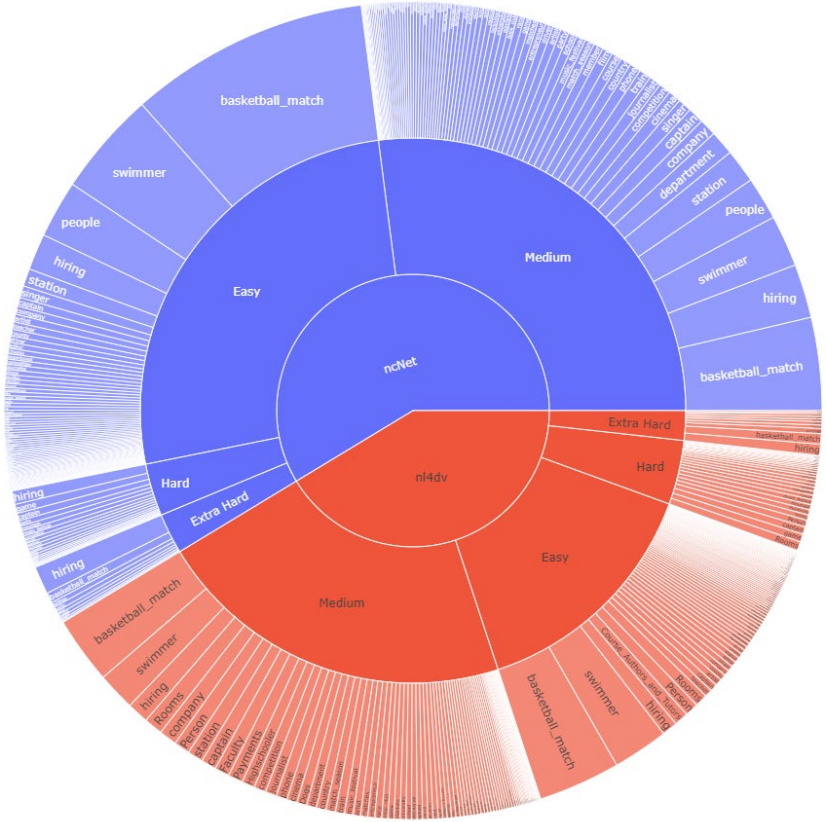


benchmark

$SSIM(V_1, V_2)$

0.898257440028004

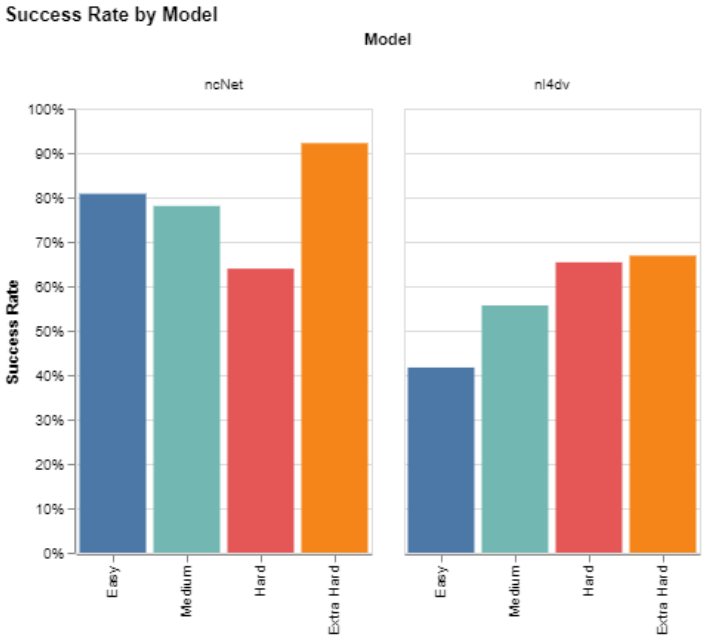
Results

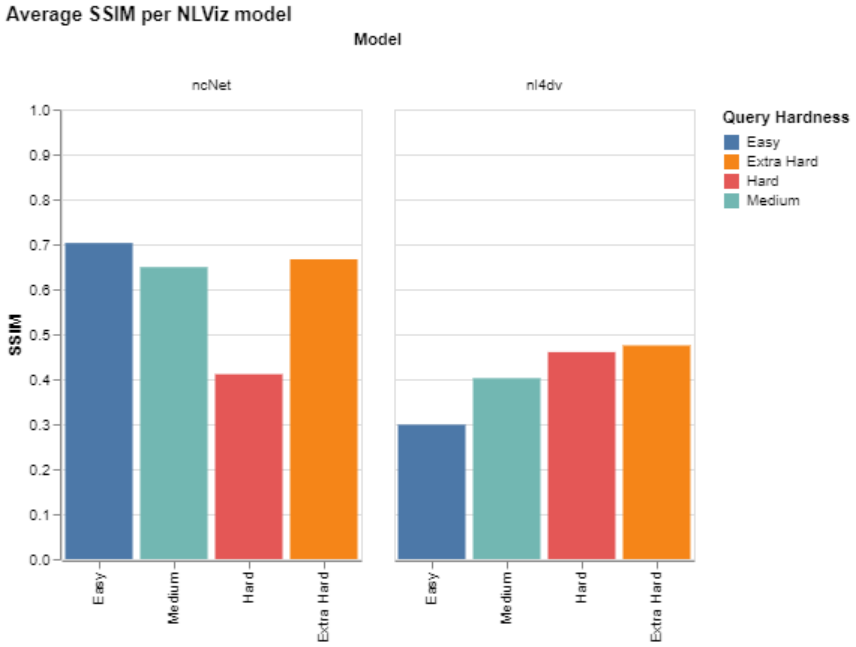


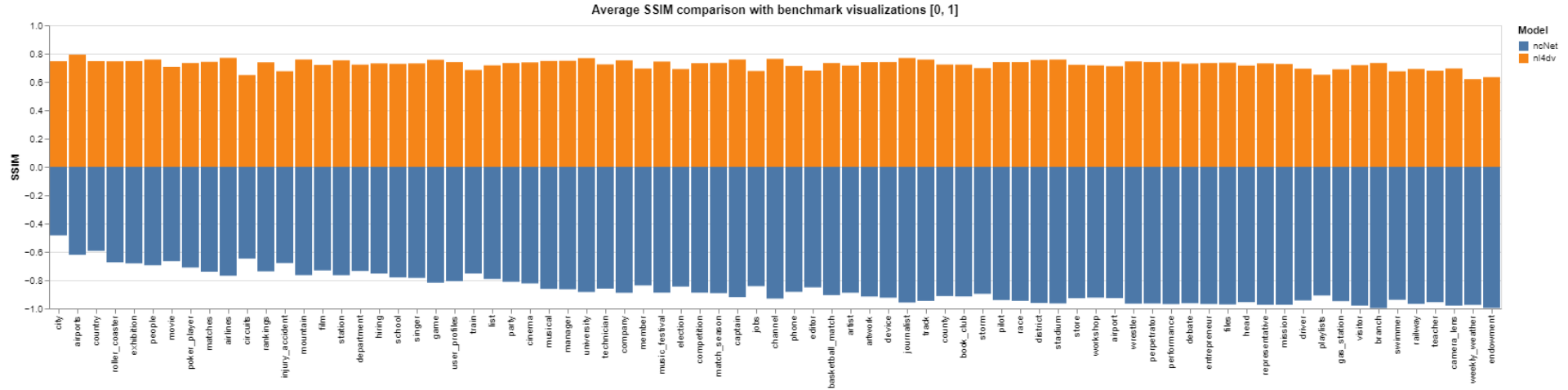
18,490 queries across 2 NL2Viz models

4 “difficulty” categories

141 unique datasets


Success Rate by Model
Model
ncNet
nl4dv
Success Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Easy
Medium
Hard
Extra Hard


Average SSIM per NLViz model
Model
ncNet
nl4dv
SSIM
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Easy
Medium
Hard
Extra Hard
Query Hardness
Easy
Extra Hard
Hard
Medium


Average SSIM comparison with benchmark visualizations [0, 1]
SSIM
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
Model
ncNet
nl4dv
city
airports
country
roller_coaster
exhibition
people
movie
poker_player
matches
airlines
circuits
rankings
injury_accident
mountain
film
station
department
hiring
school
singer
game
user_profiles
train
list
party
cinema
musical
manager
university
technician
company
member
music_festival
election
competition
match_season
captain
jobs
channel
phone
editor
basketball_match
artist
artwork
device
journalist
track
county
book_club
storm
pilot
race
district
stadium
store
workshop
airport
wrestler
perpetrator
performance
debate
entrepreneur
files
head
representative
mission
driver
playlists
gas_station
visitor
branch
swimmer
railway
teacher
camera_lens
weekly_weather
endowment

Next Steps

- Find/combine more metrics for model evaluation, or come up with a new one
- Evaluate more NL2Viz models using existing pipeline
- Online tool – generate SSIM metrics on-demand
- Online tool – built-in specification editor

Thank you!

Image References

- https://logos-world.net/wp-content/uploads/2021/10/Tableau-Emblem.png
- https://user-images.githubusercontent.com/315810/92254613-279c8000-ee9f-11ea-9b73-5622a7d95f3f.png
- https://upload.wikimedia.org/wikipedia/commons/thumb/5/58/Vega-Lite_Logo.svg/2560px-Vega-Lite_Logo.svg.png