

Data Visualization

The importance of principled design

The need for sound principles

Visualization is intuitive!

I already know how to do it and my audience already understands how to interpret it.

OK.

So what color
is this dress?



Visual perception might be complicated.

So is the science behind getting it right.



Ο Κεφ' αριστερά
αριστερά
αριστερά
αριστερά
αριστερά

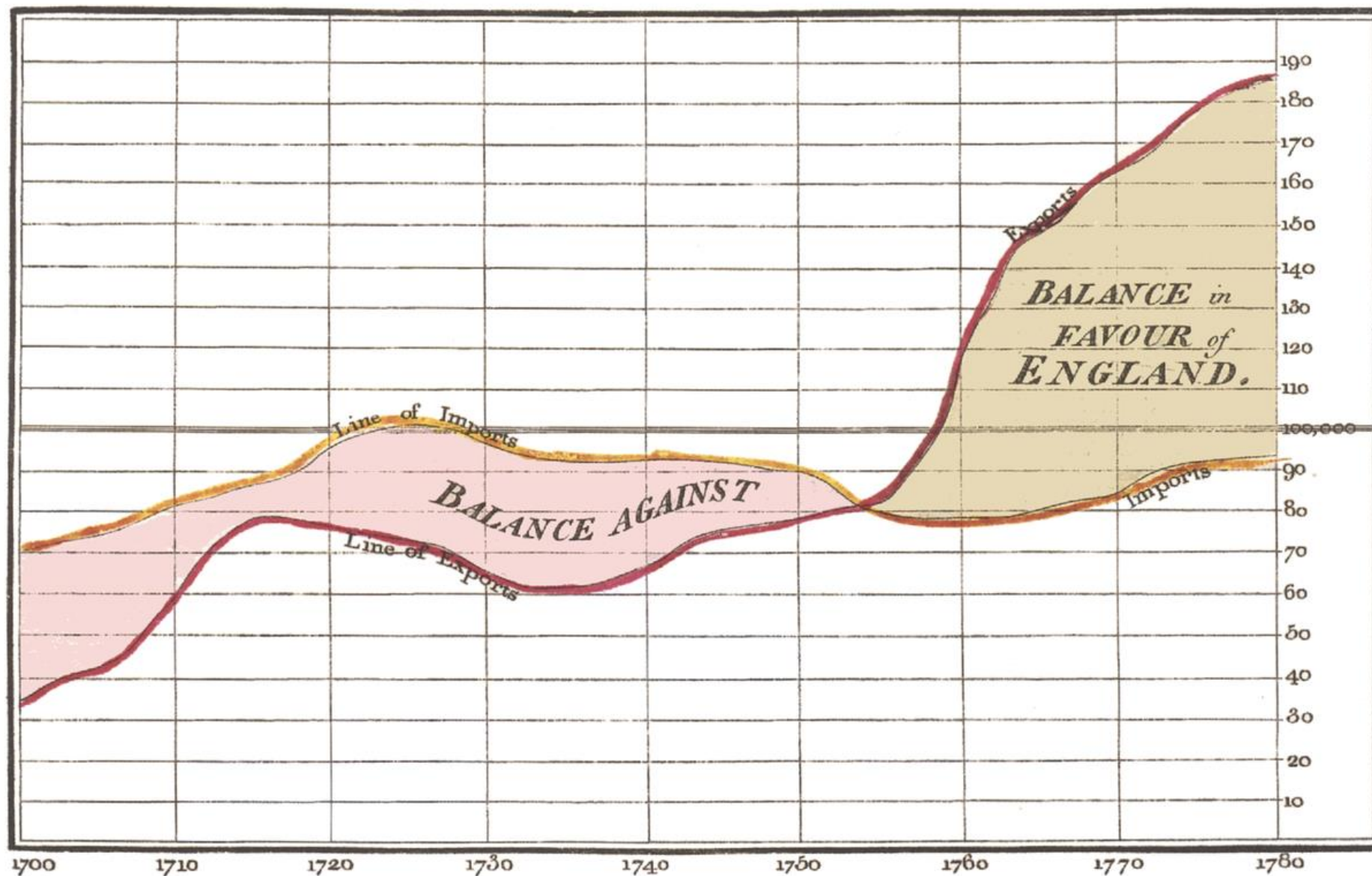
Ο Κεφ' αριστερά
αριστερά
αριστερά
αριστερά

Ο Κεφ' αριστερά
αριστερά
αριστερά
αριστερά

Ο Κεφ' αριστερά
αριστερά
αριστερά
αριστερά

Ο Κεφ' αριστερά
αριστερά
αριστερά
αριστερά

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



The Bottom line is divided into Years, the Right hand line into £10,000 each.

How Charts Lie



Getting Smarter about
Visual Information

Alberto Cairo



*If a picture says a thousand words,
that just means it can lie at scale.*

Principles

Keys to Making Great Visualizations:

What is the Message!

Know How the Brain Processes

Declutter

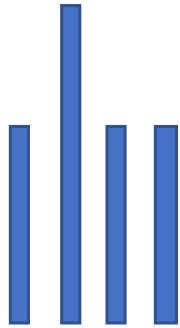
Focus Attention

Ask Yourself Throughout:

“What is the Message?”

Visual Processing Speed: 'Pre-attentive Attributes'

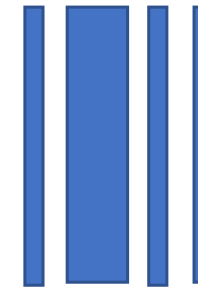
Length



2D Position



Width



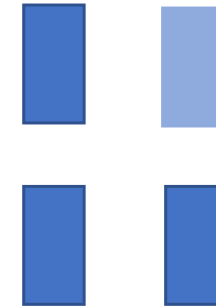
Shape



Color (hue)



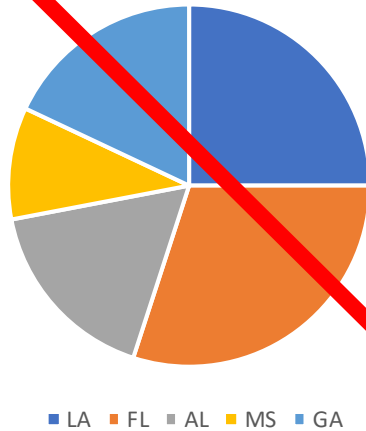
Color (intensity)



See also: Gestalt Principles (Proximity, Similarity, Enclosure, Closure, Continuity, Connection)

Never choose:

Pie Chart



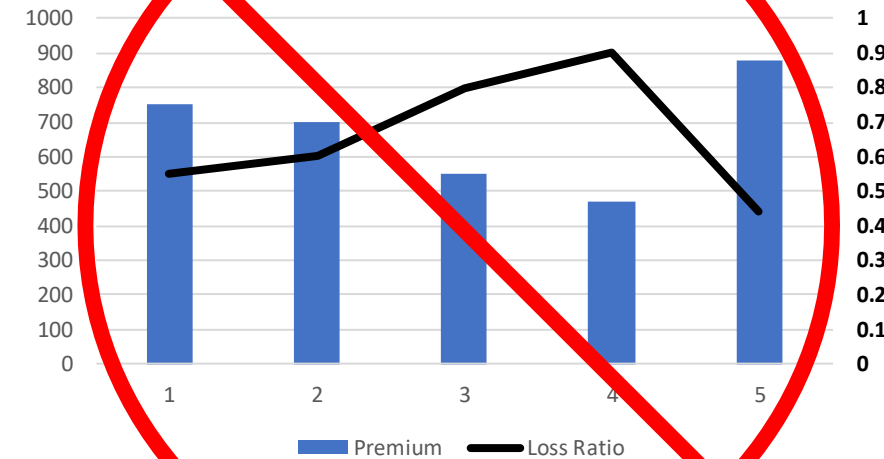
...Instead try:
(stacked) bar graph for shares

3D graph



...Instead try:
2-dimensional only

Double-Y Axis



...Instead try:
Graph bar and line separately

Declutter:

Use the 'ink' for data, not non-data!

No gridlines

No border

Fewer x-axis and y-axis tickmarkers

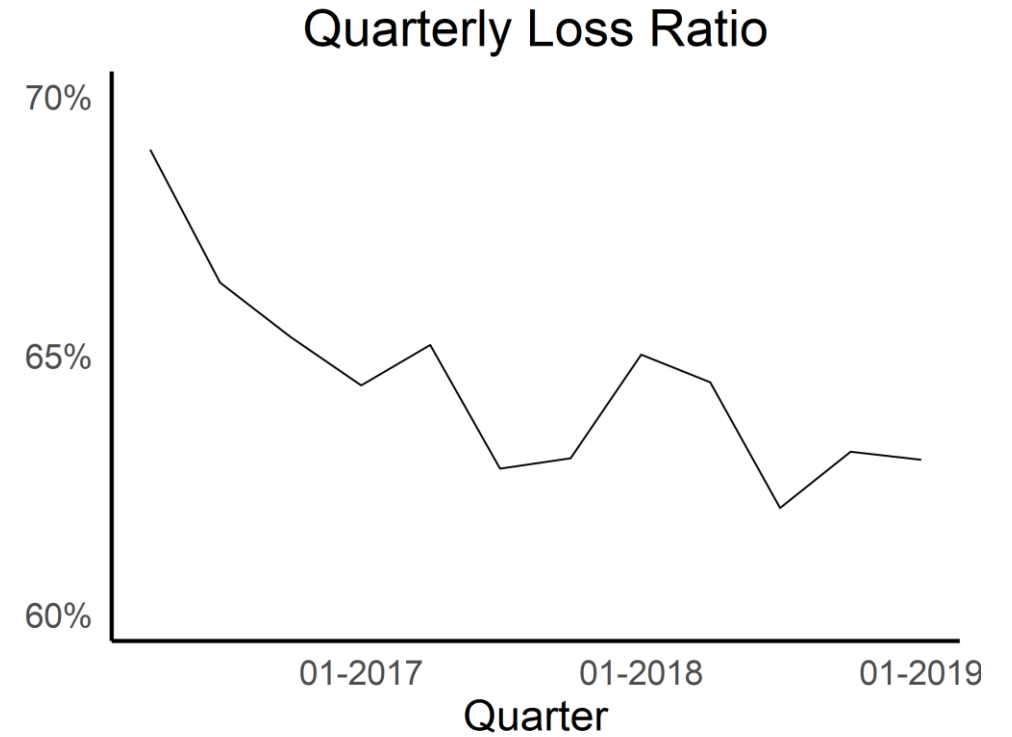
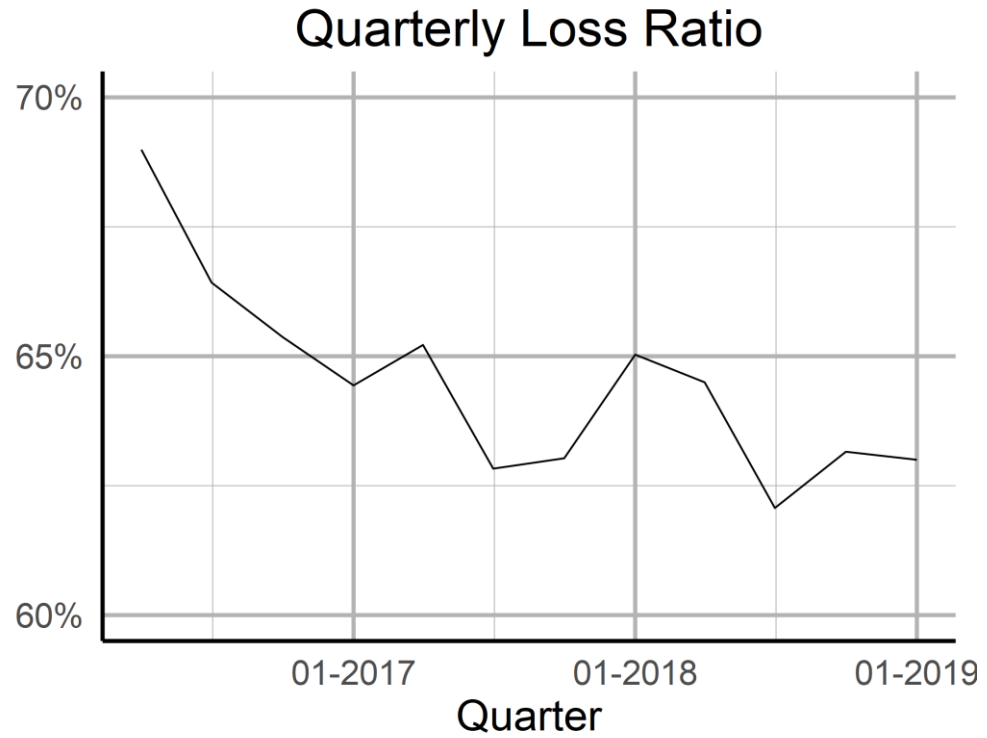
Axis labels – streamline text, and no 90-degree orientation for Y label

No legend – rather, directly label data within graph, with color coding to data

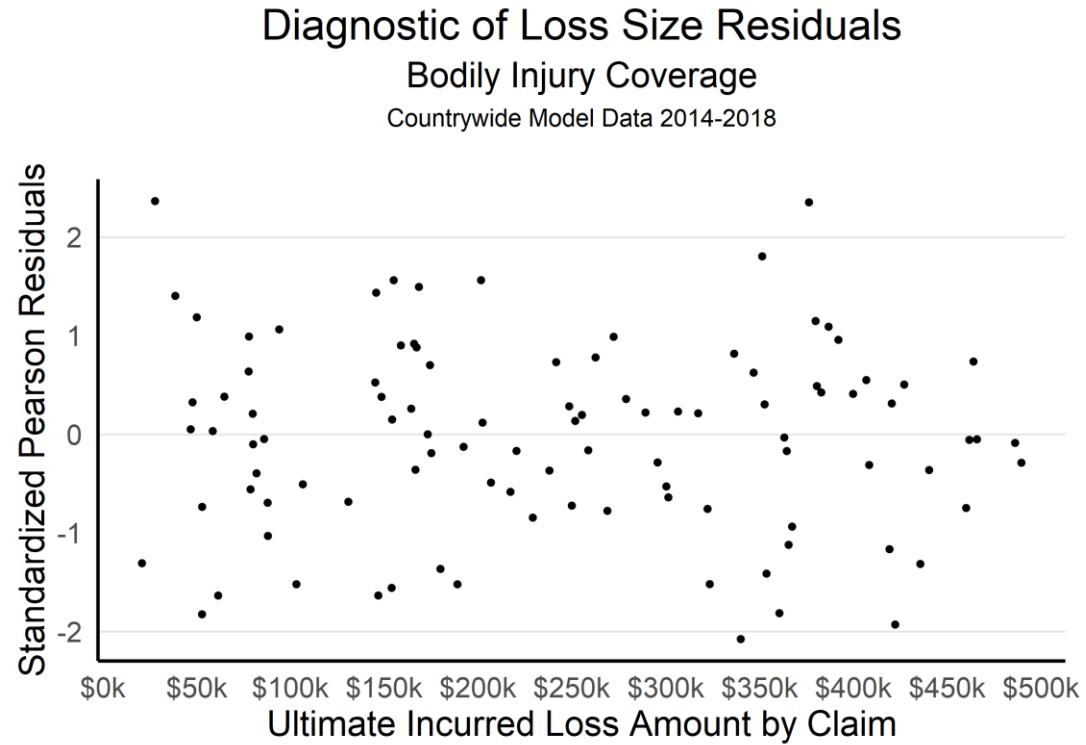
No Subheaders (e.g. precise nature of data or yrs used etc. moved to footnote)

Gray out axis lines and other chart-junk; ok to gray less important data

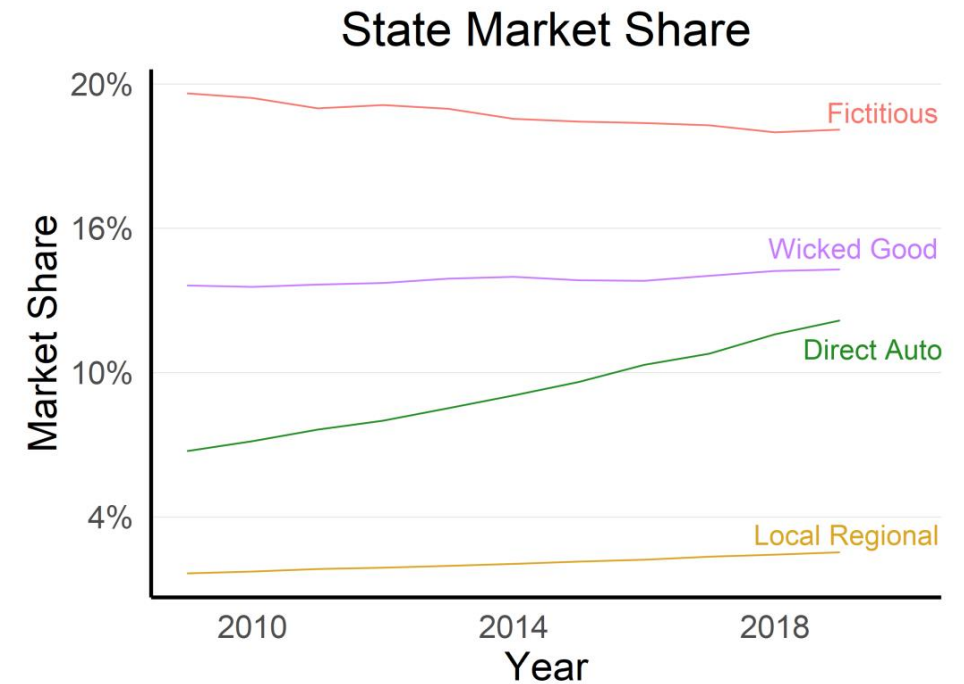
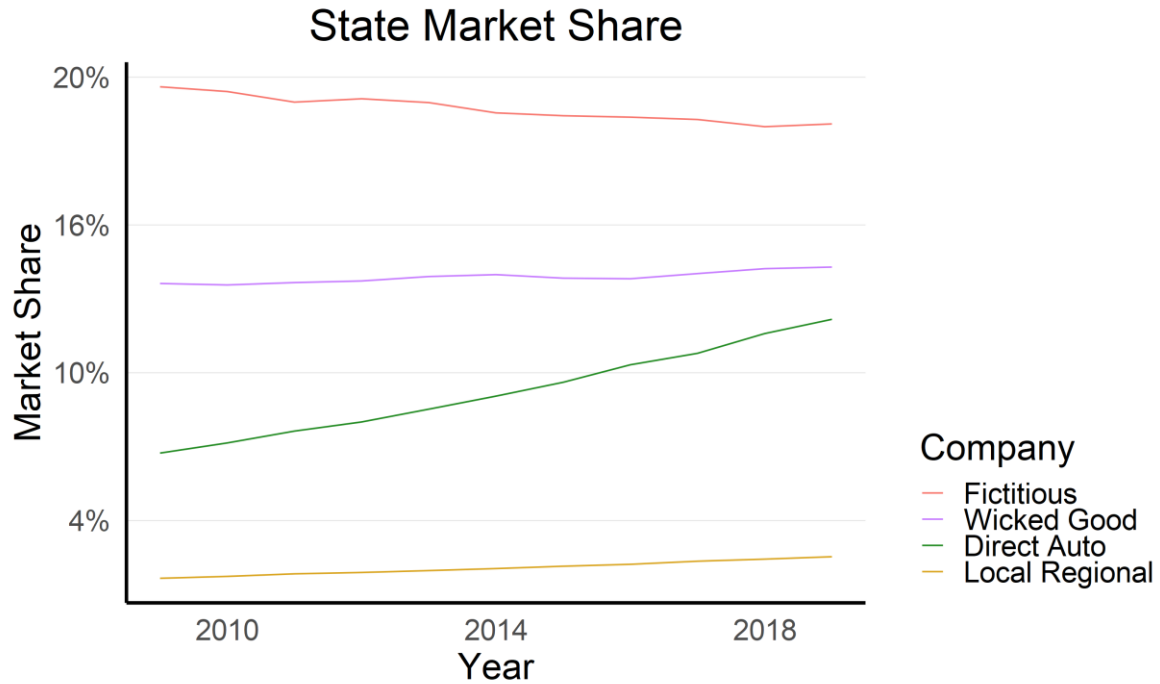
Declutter: Remove Gridlines



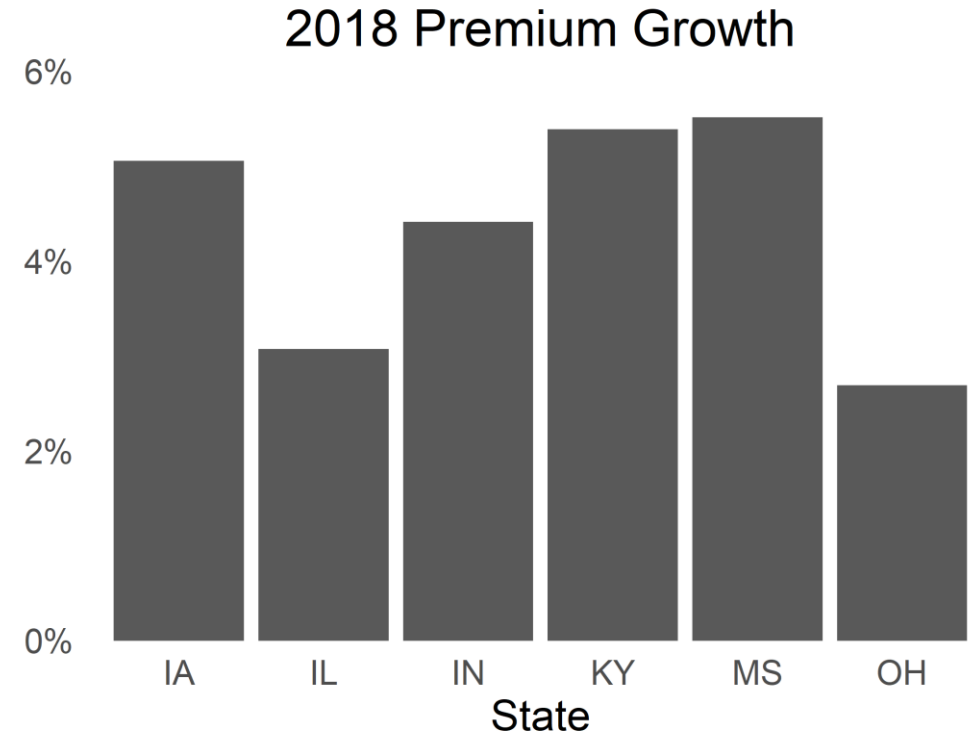
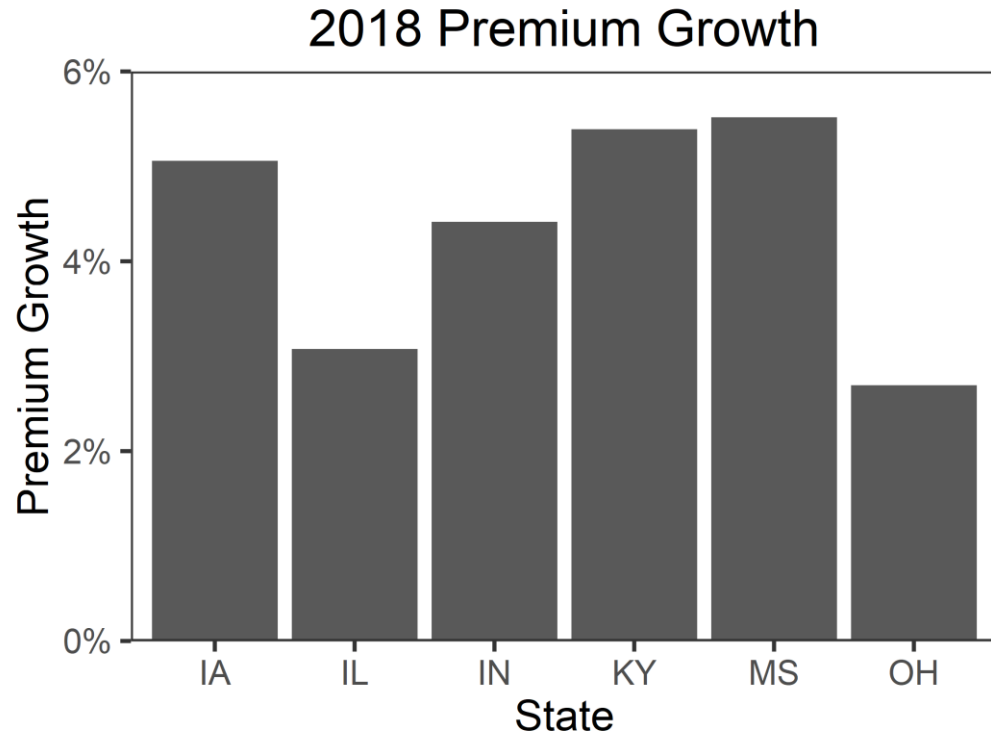
Declutter: Reduce Text



Declutter: Legends



Declutter: Remove axis ticks/graph border



Focus Attention, Text Version:

None: Direct Auto is increasing market share the fastest!

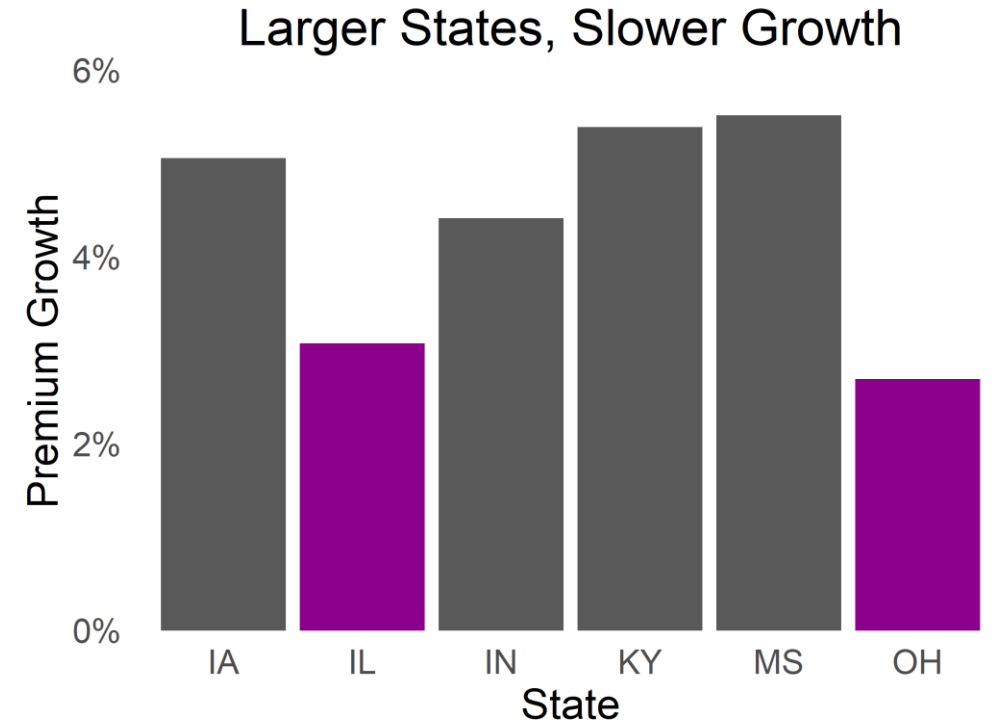
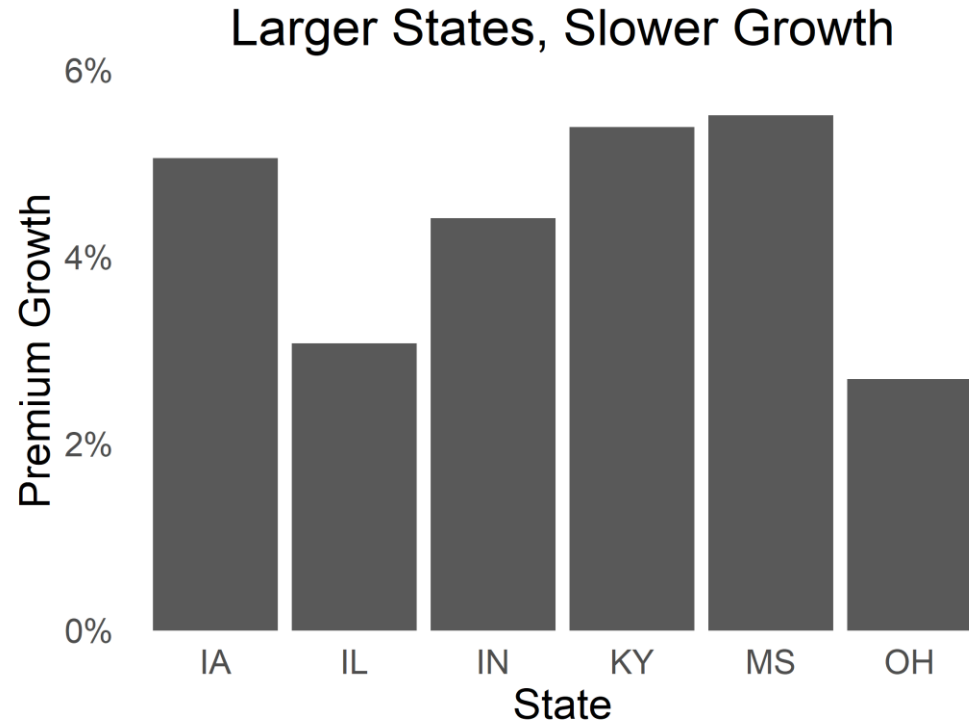
Size: Direct Auto is increasing market share the fastest!

Bold: **Direct Auto** is **increasing** market share the **fastest**!

Color: Direct Auto is increasing market share the fastest!

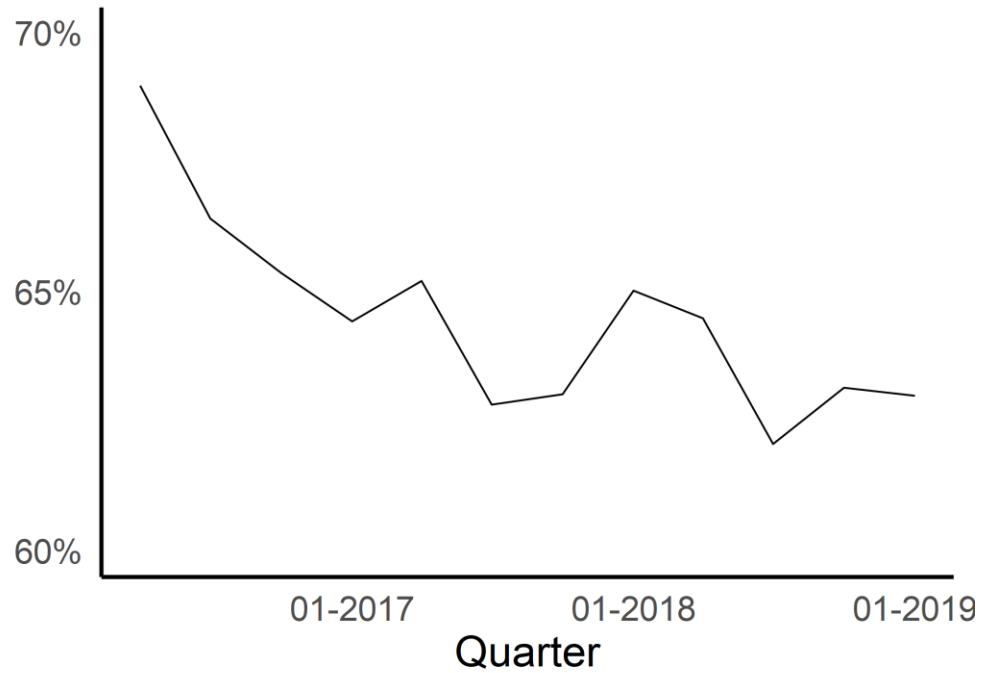
Gray: Direct Auto is increasing market share the fastest!

Focus Attention: Color

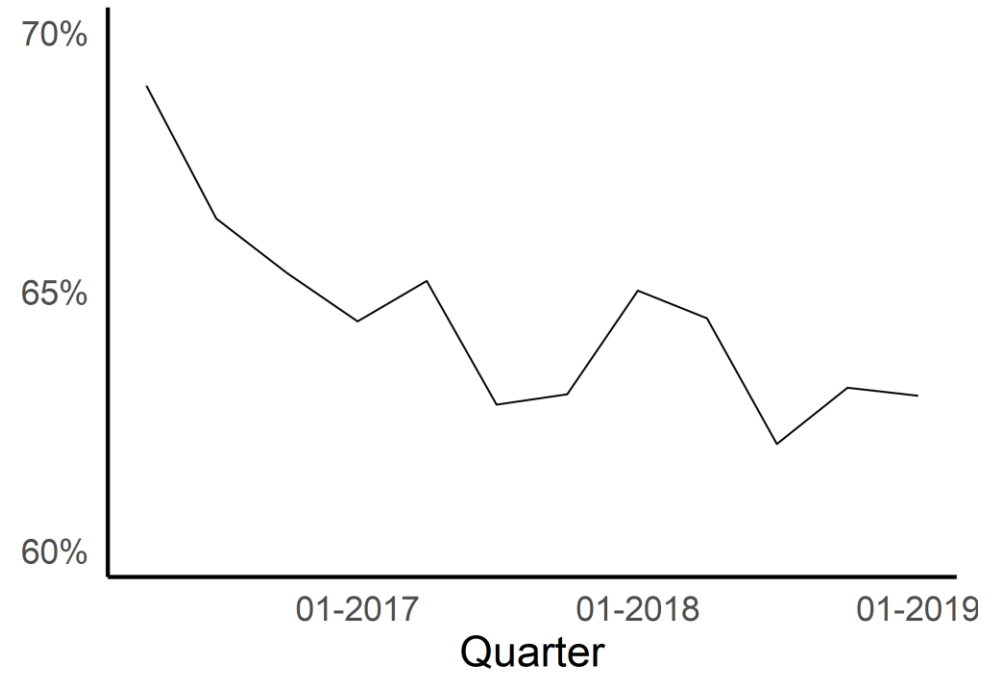


Focus Attention: "Takeaway" Titles

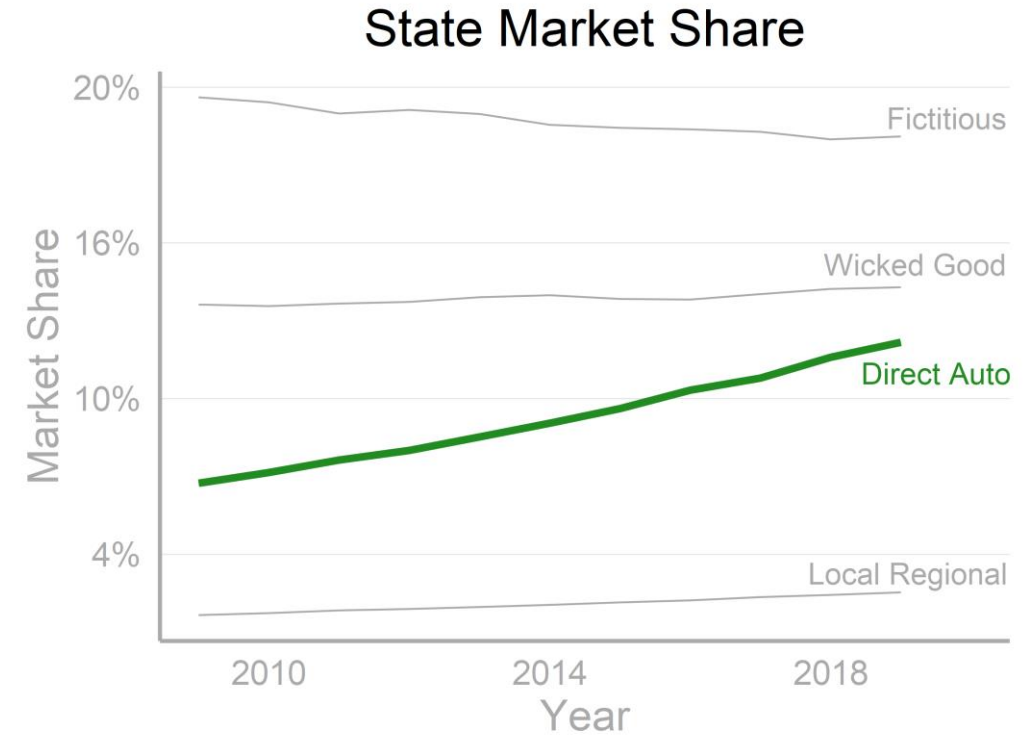
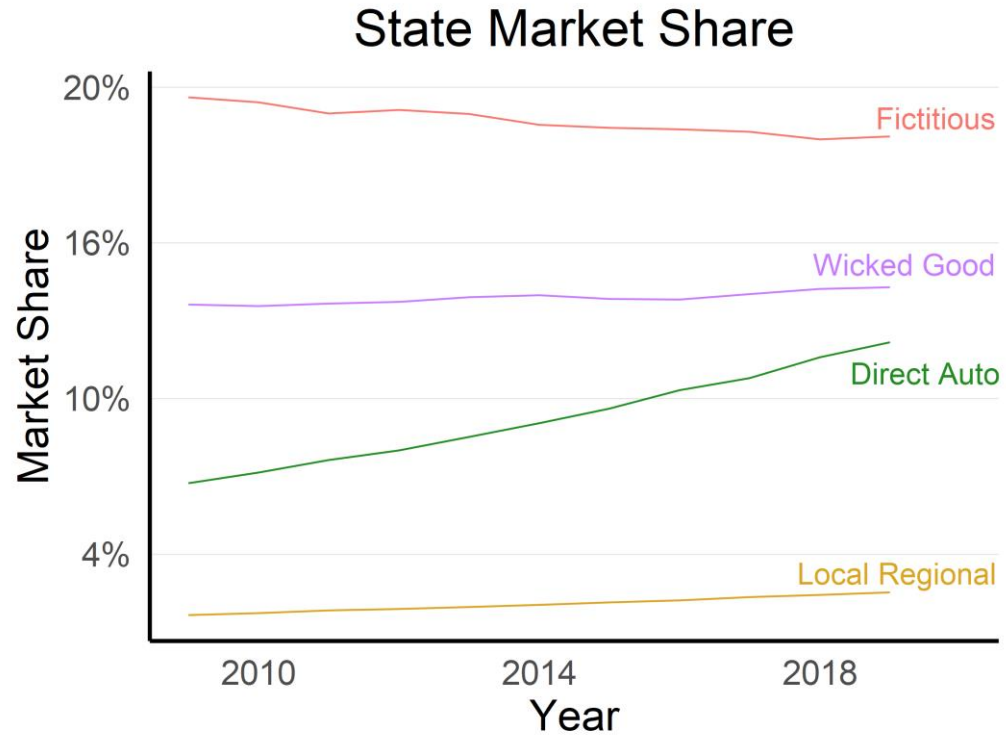
Quarterly Paid Loss Ratios 2016-2019



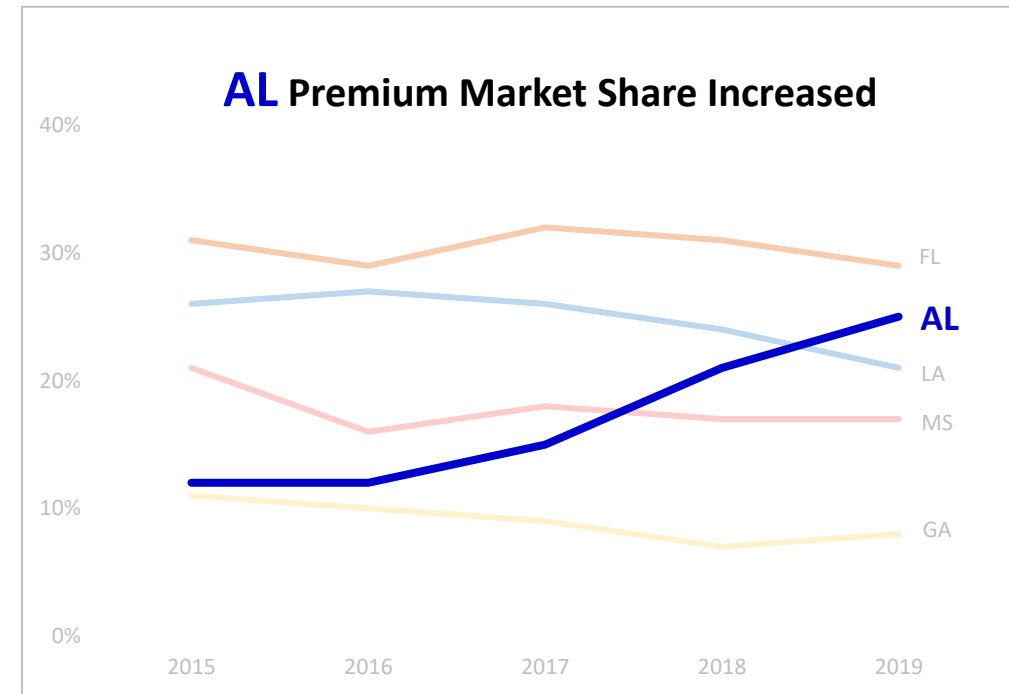
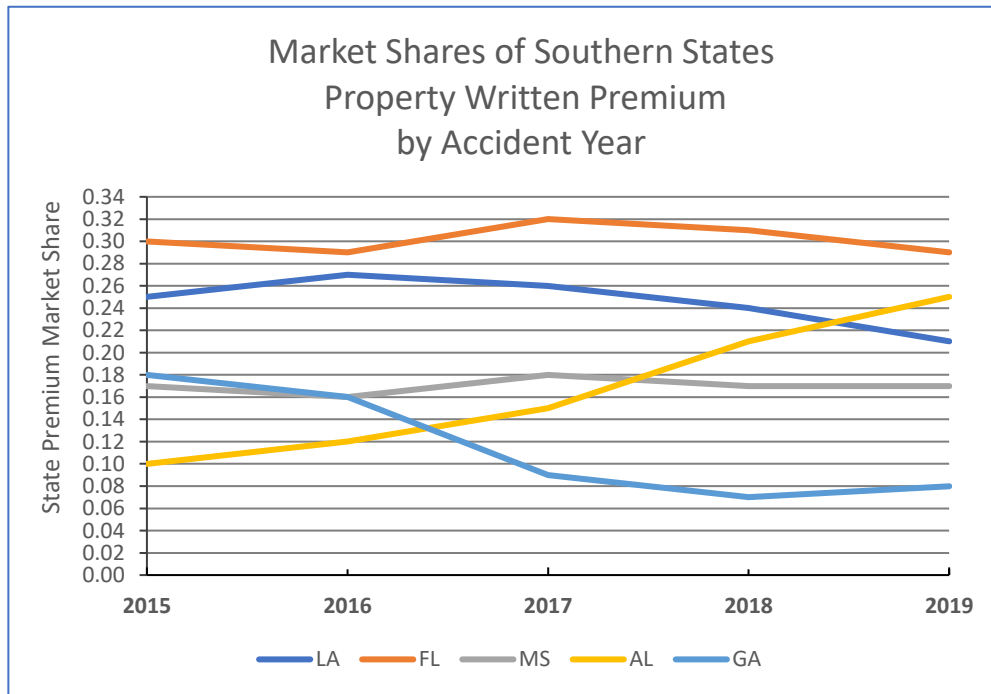
Loss Ratio Improvement Continues



Focus Attention: Gray less important content



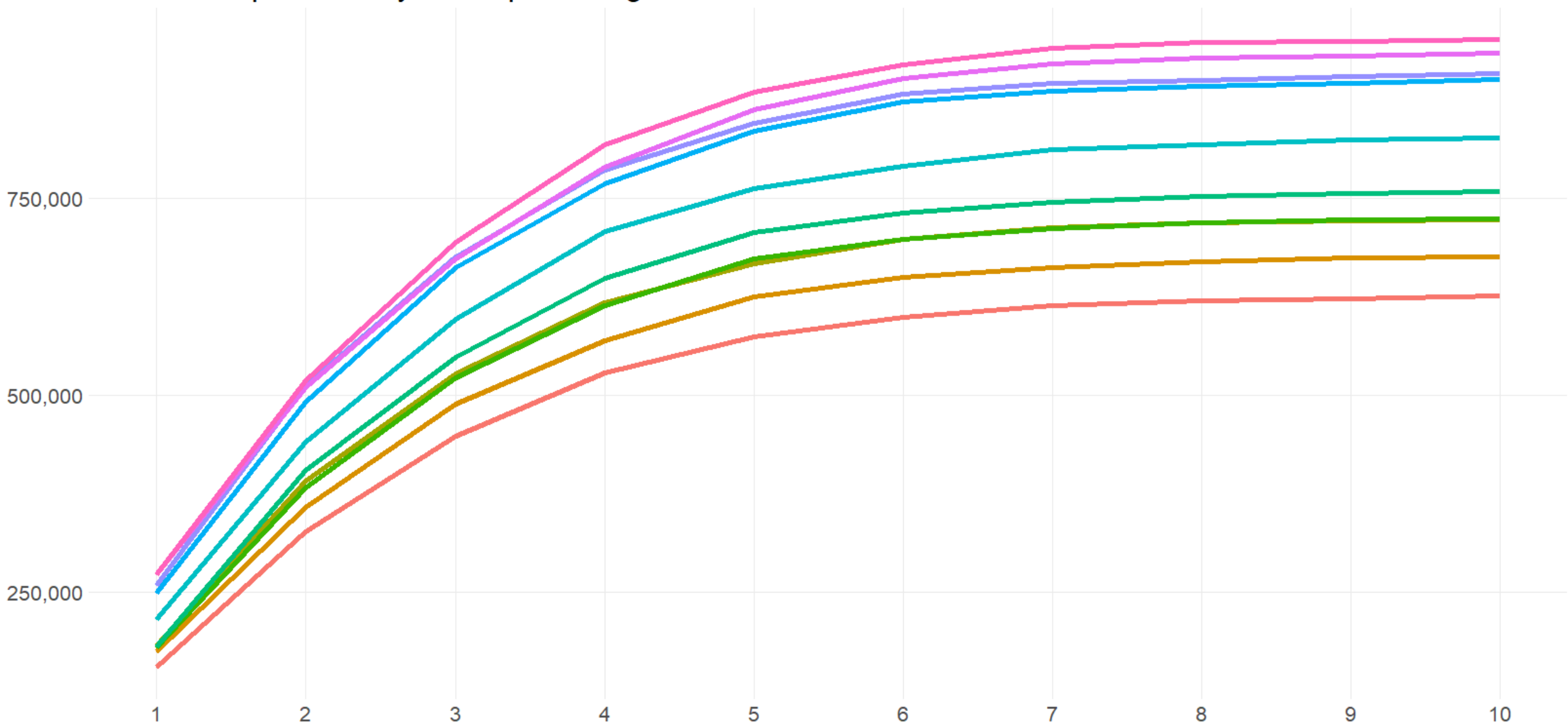
Message, Declutter, and Focus Attention:



Application 1

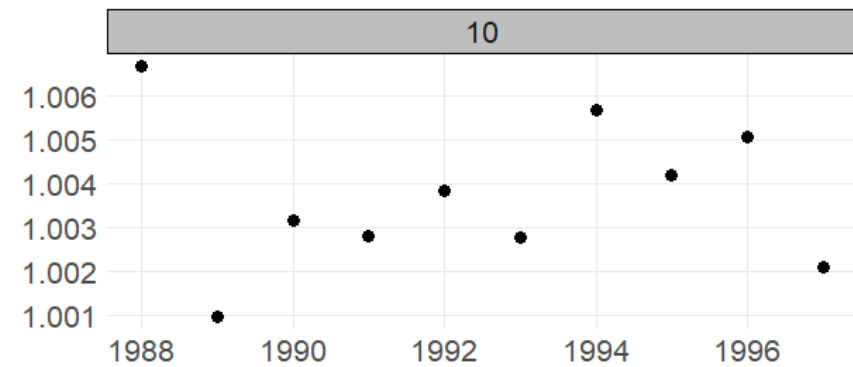
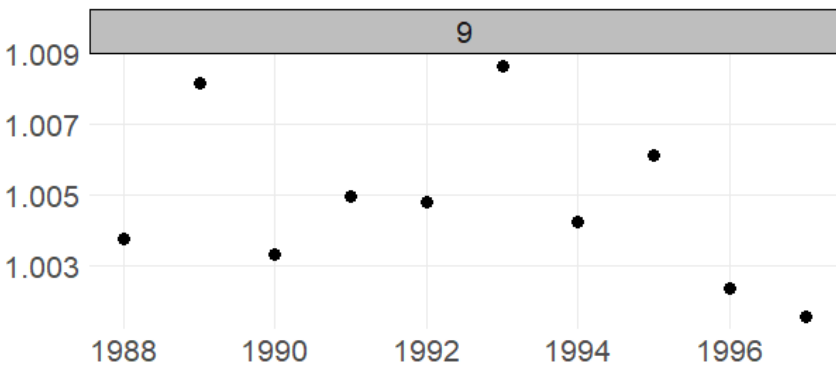
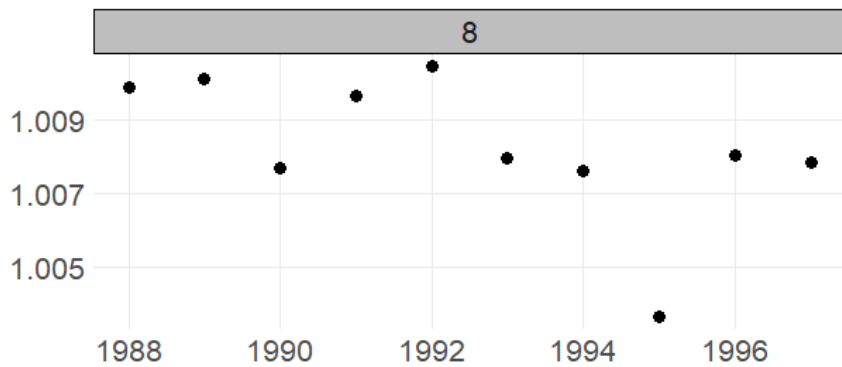
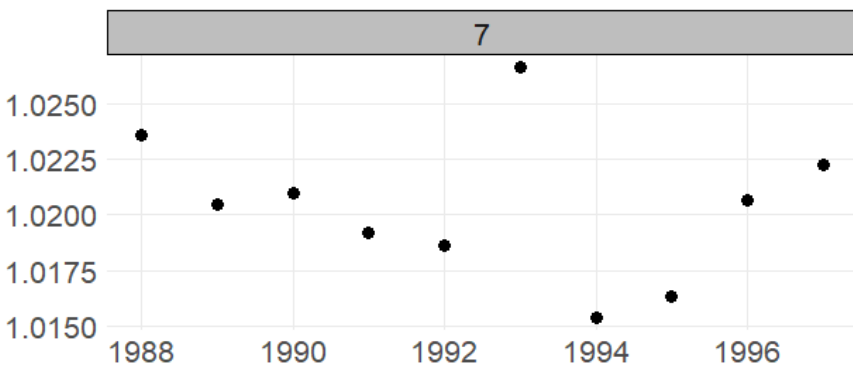
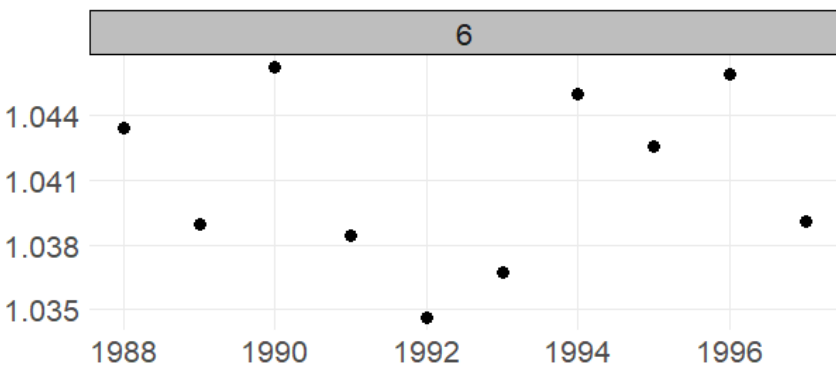
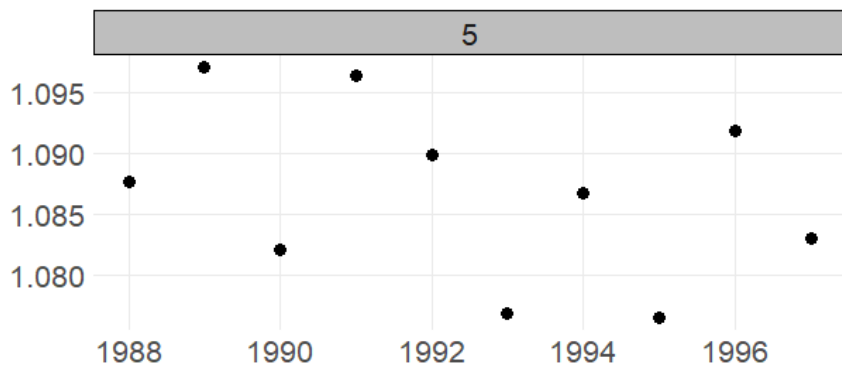
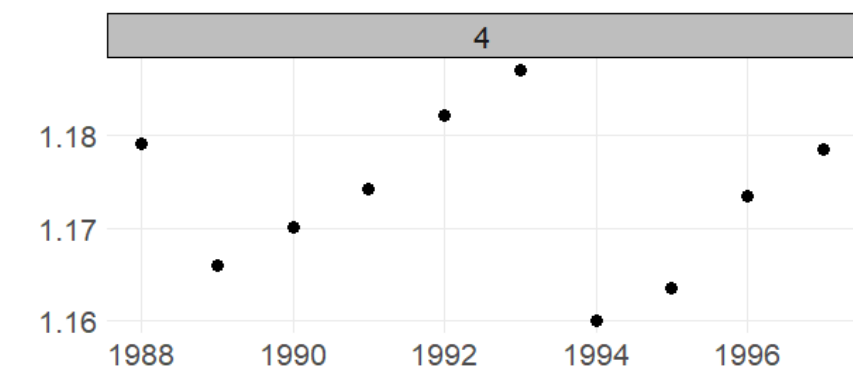
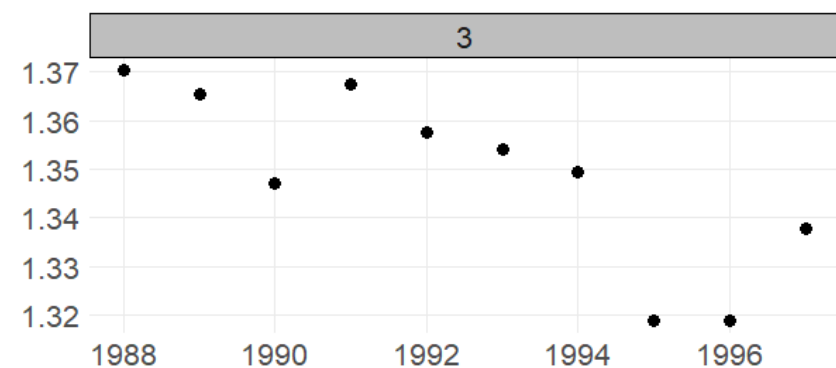
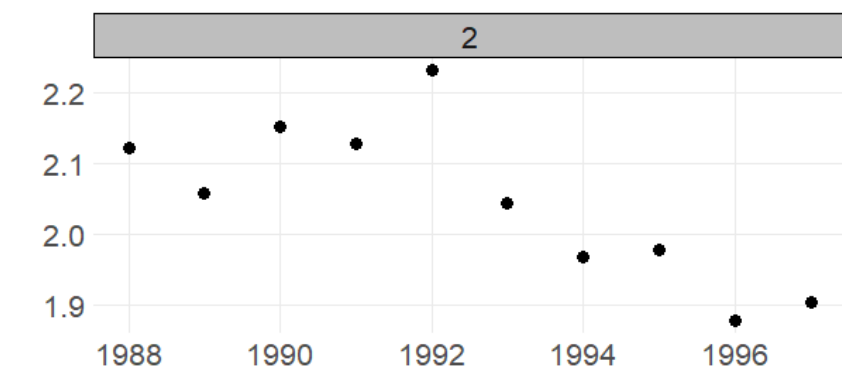
Loss Reserving

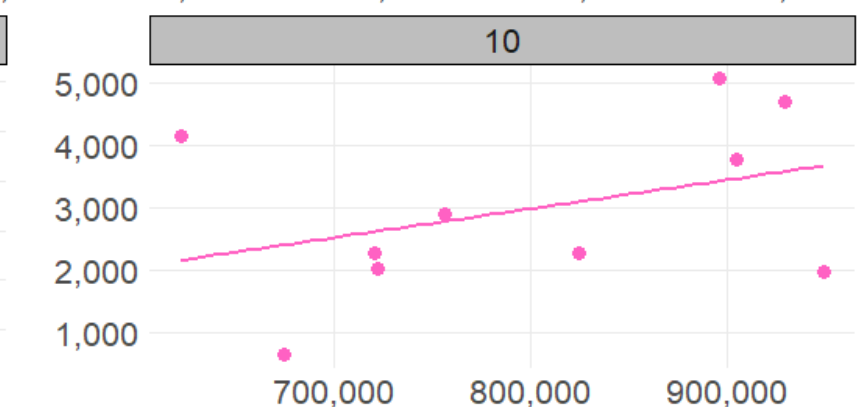
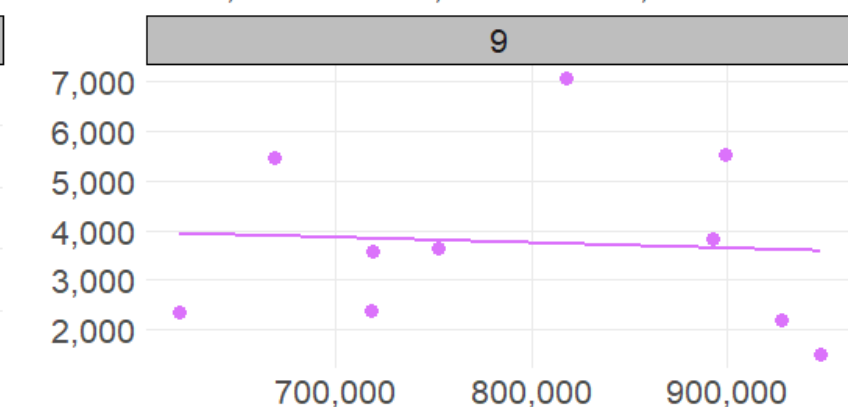
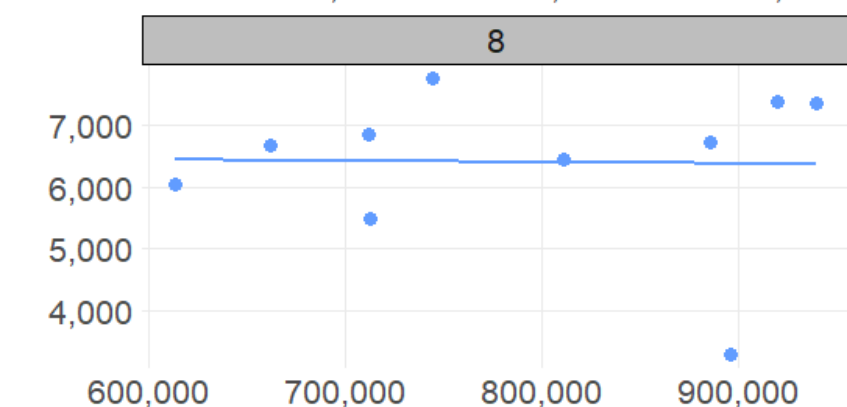
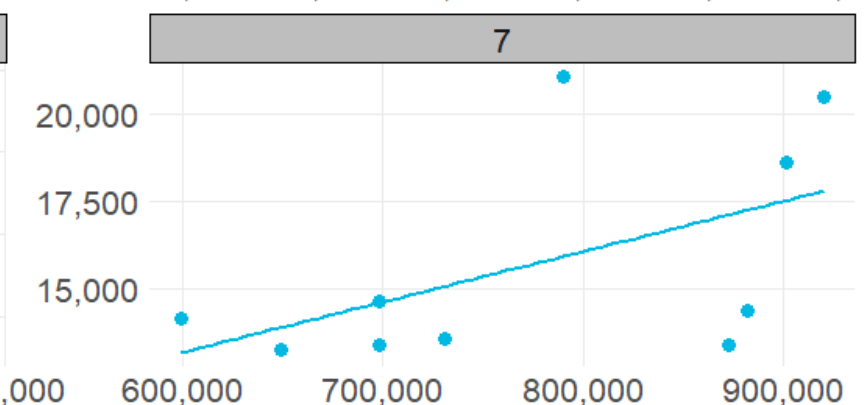
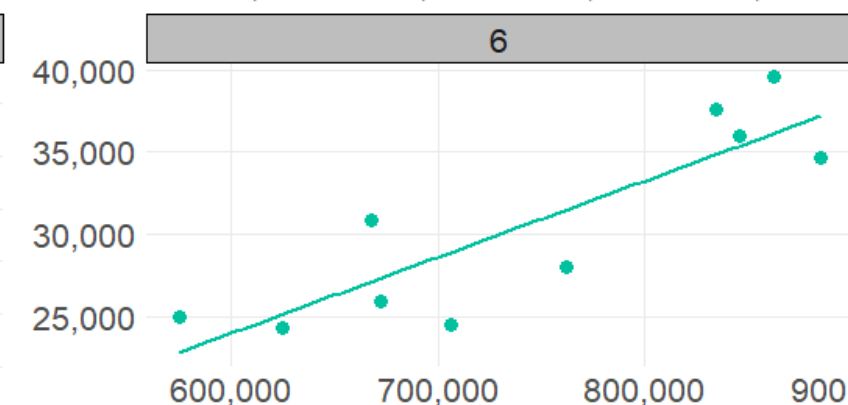
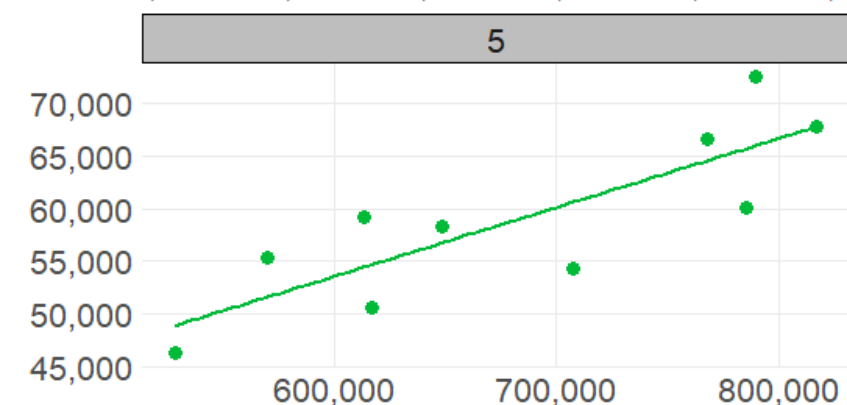
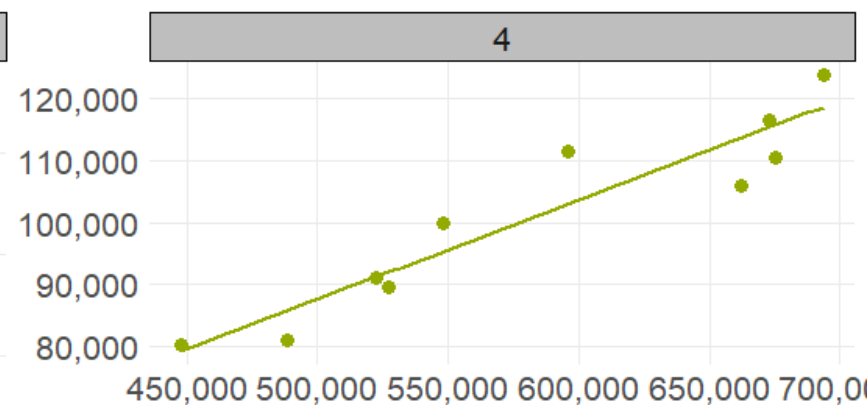
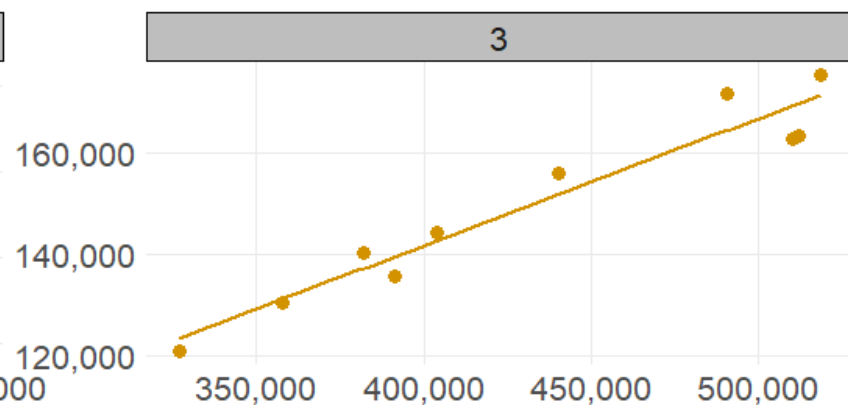
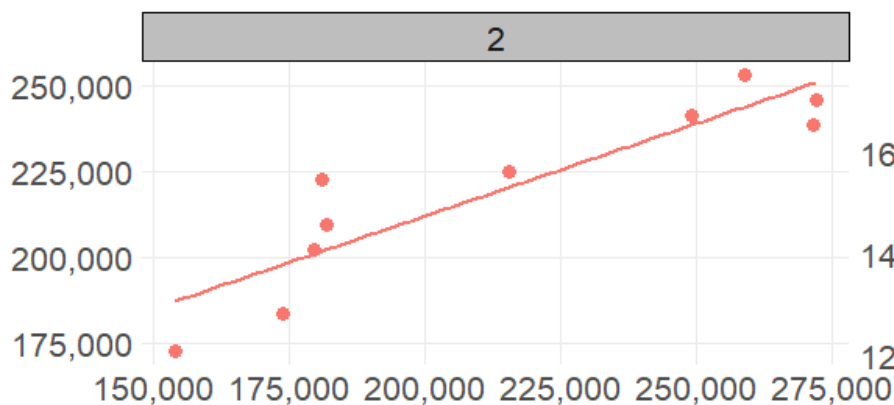
Cumulative paid loss by development lag



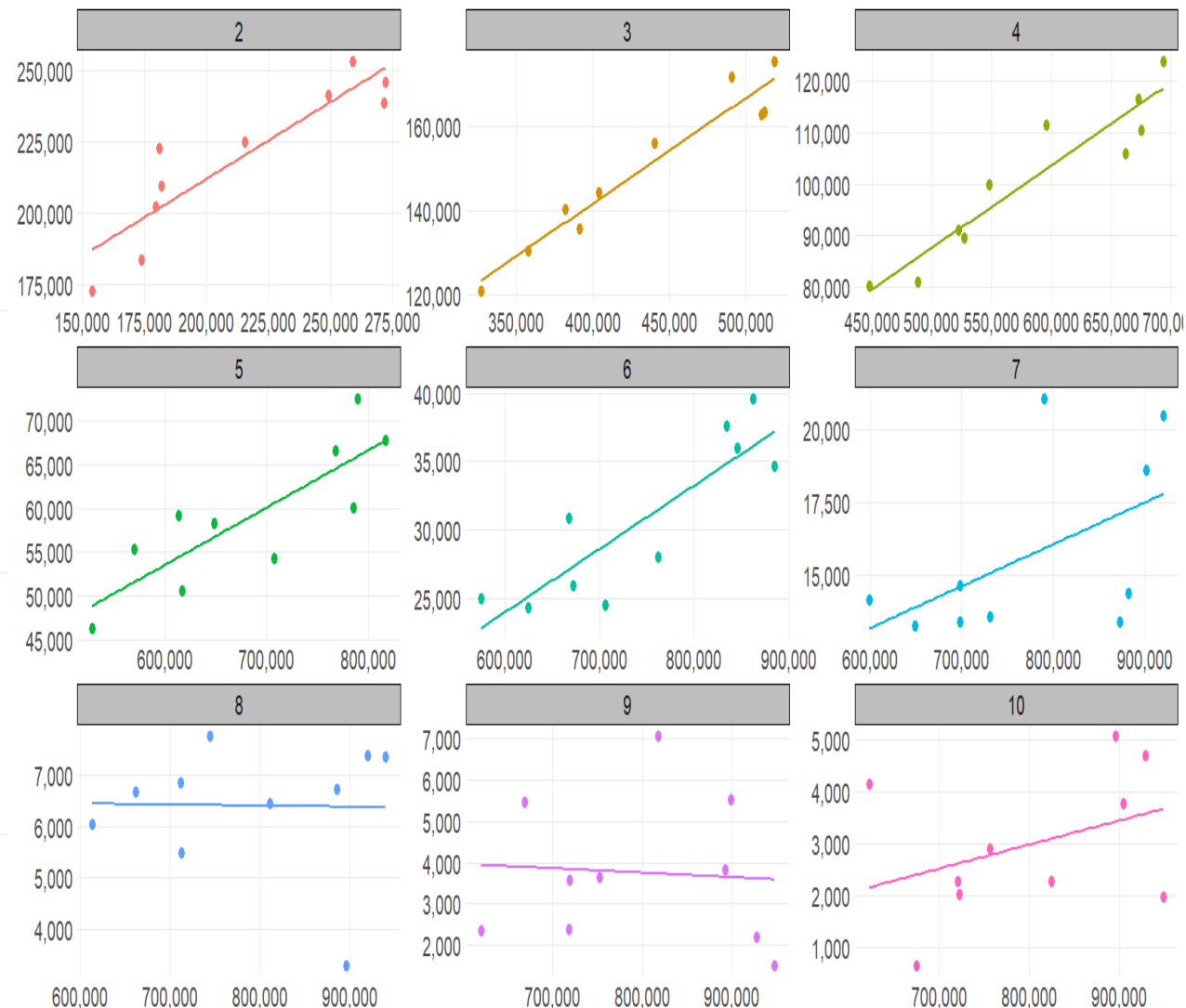
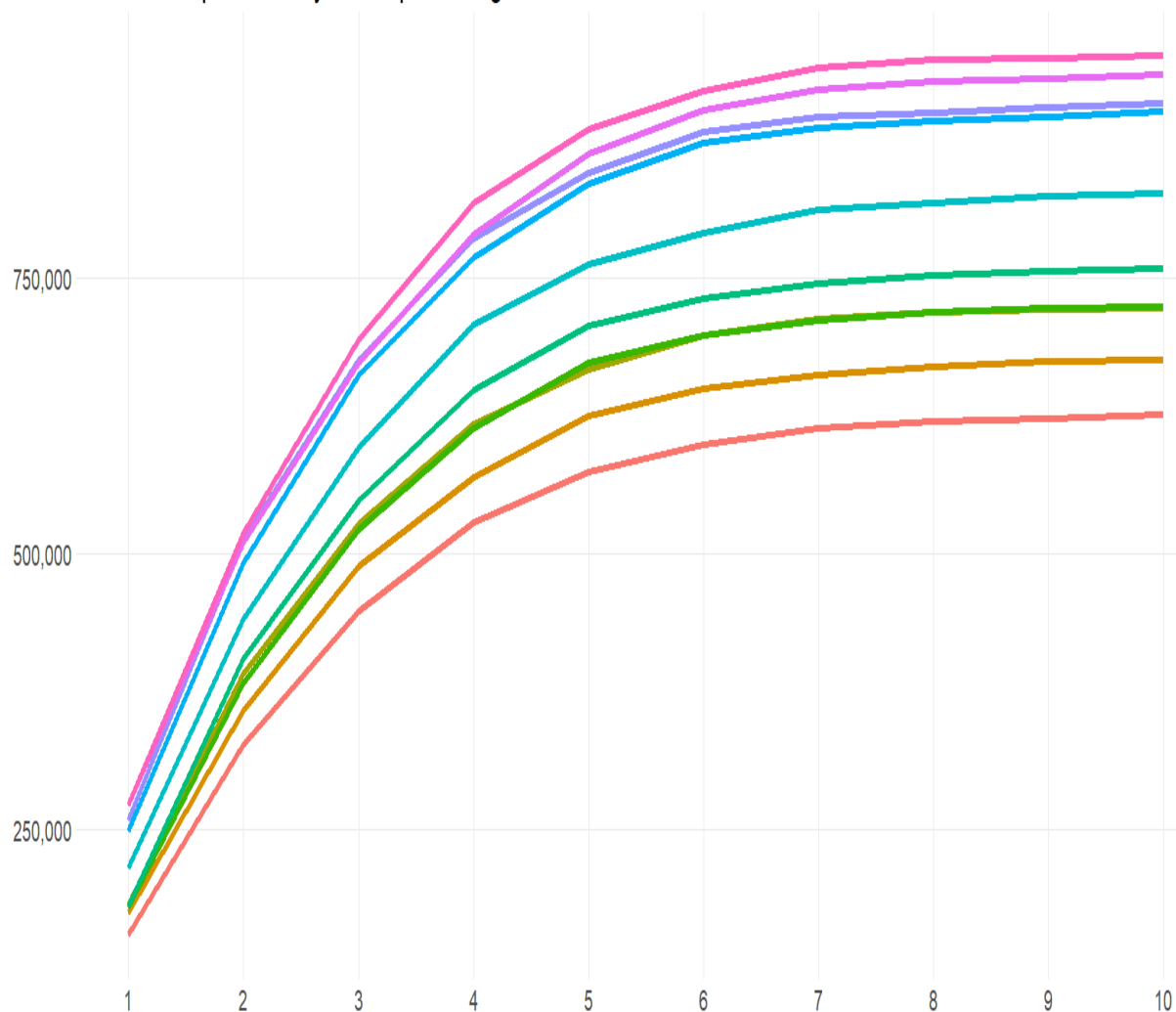
“A sure sign of a puzzle is that the graphic must be interpreted through a verbal, rather than a visual process”

- Edward Tufte





Cumulative paid loss by development lag

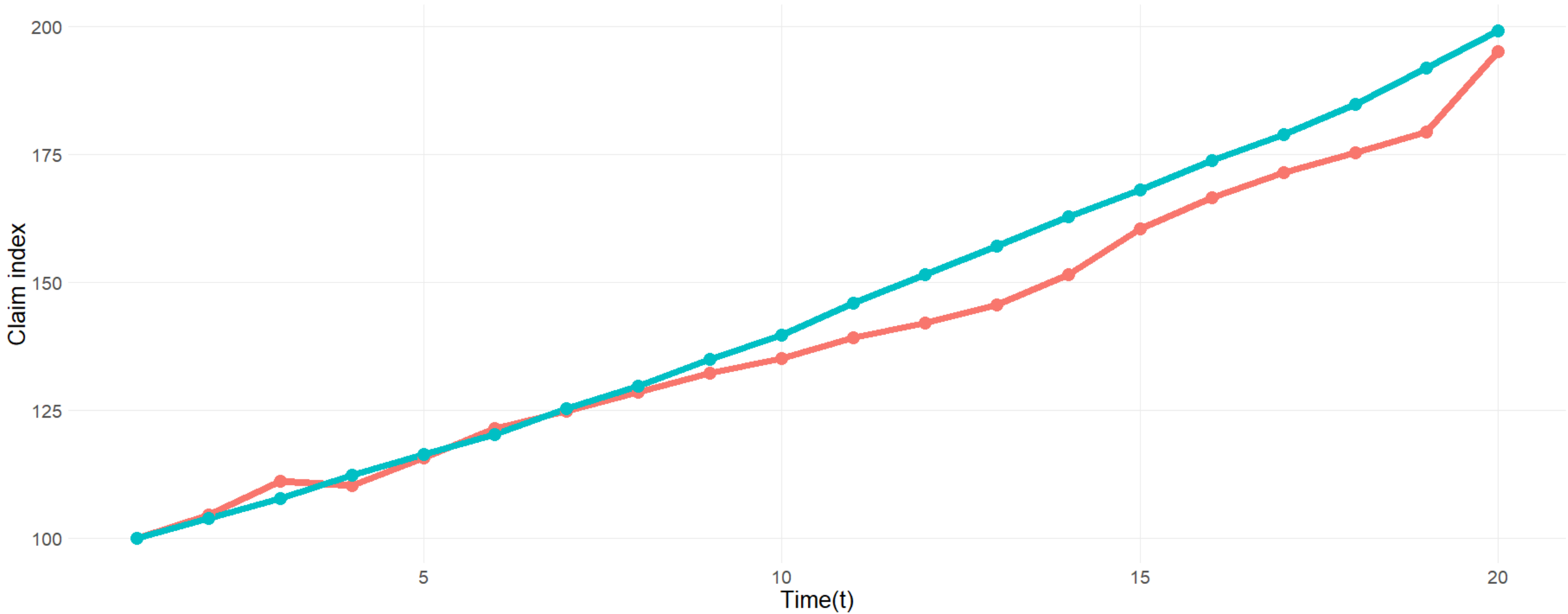


Application 2

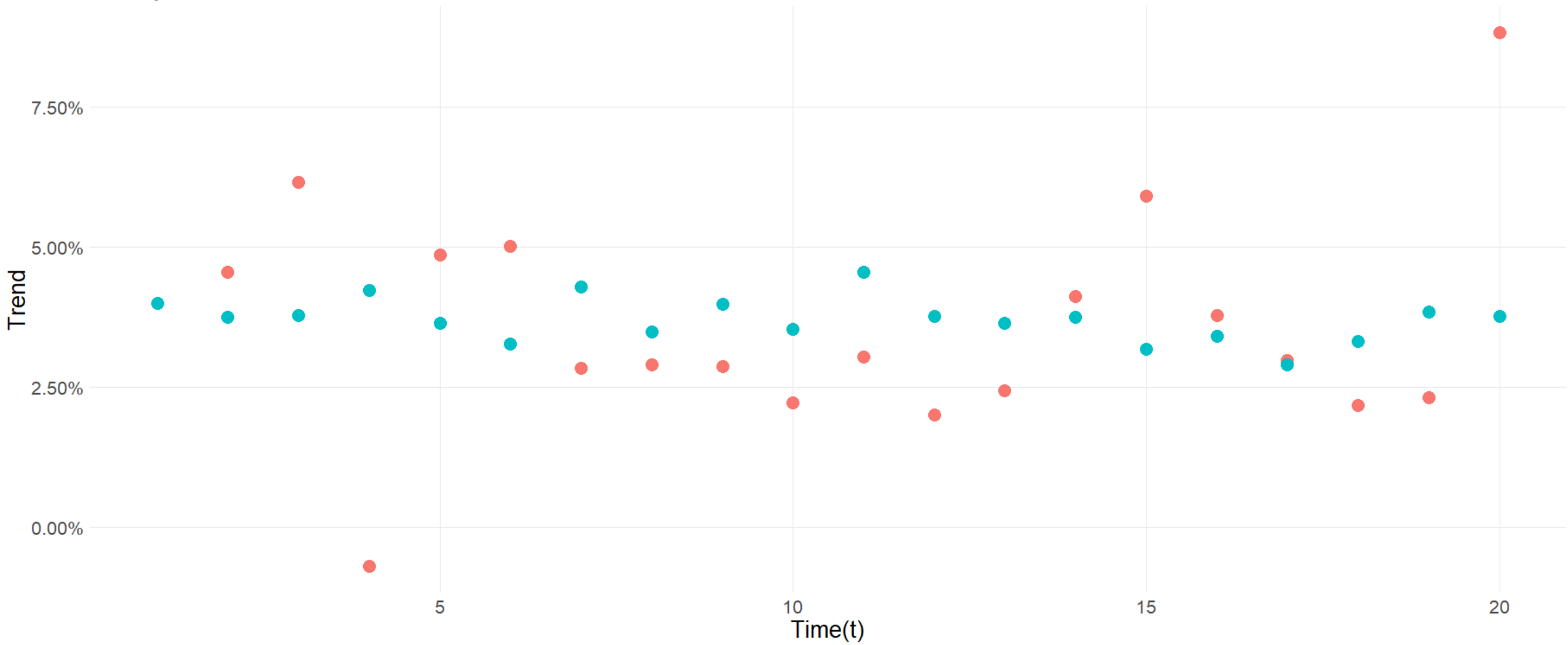
Another Example

Relative mean claim size

Mean claim indexed to value of 100 at $t = 1$

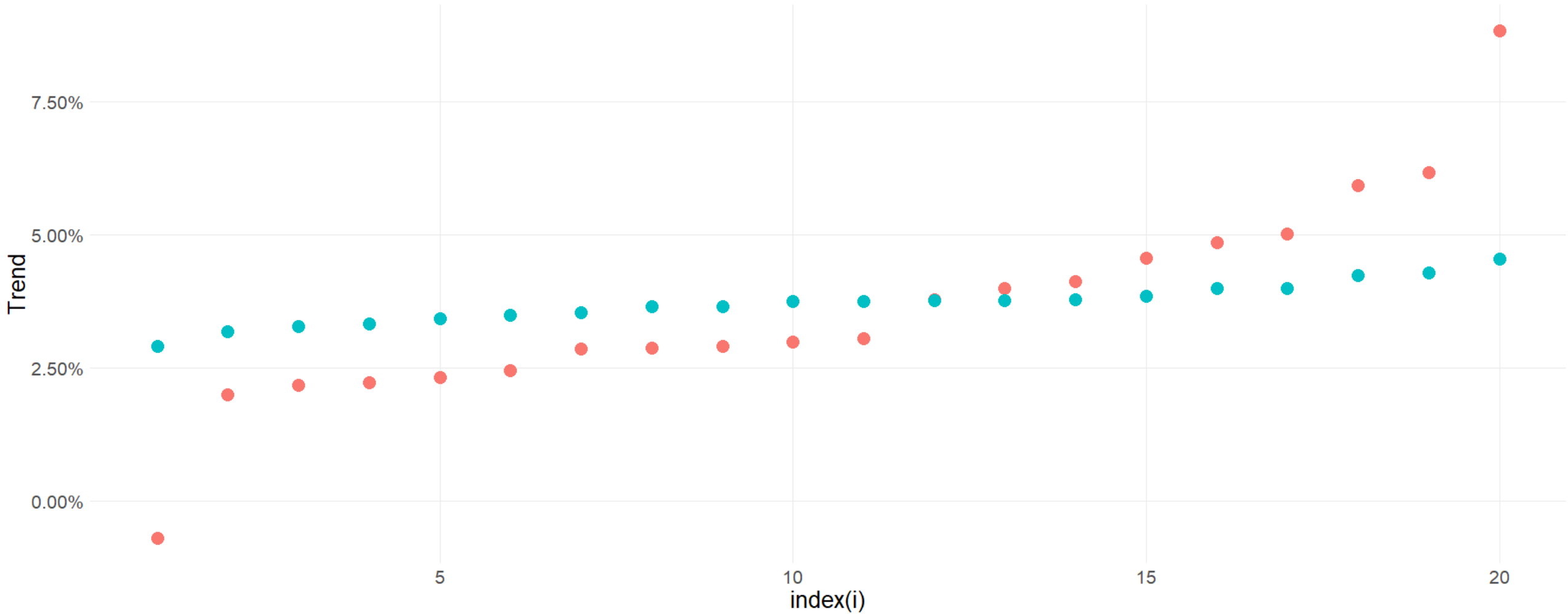


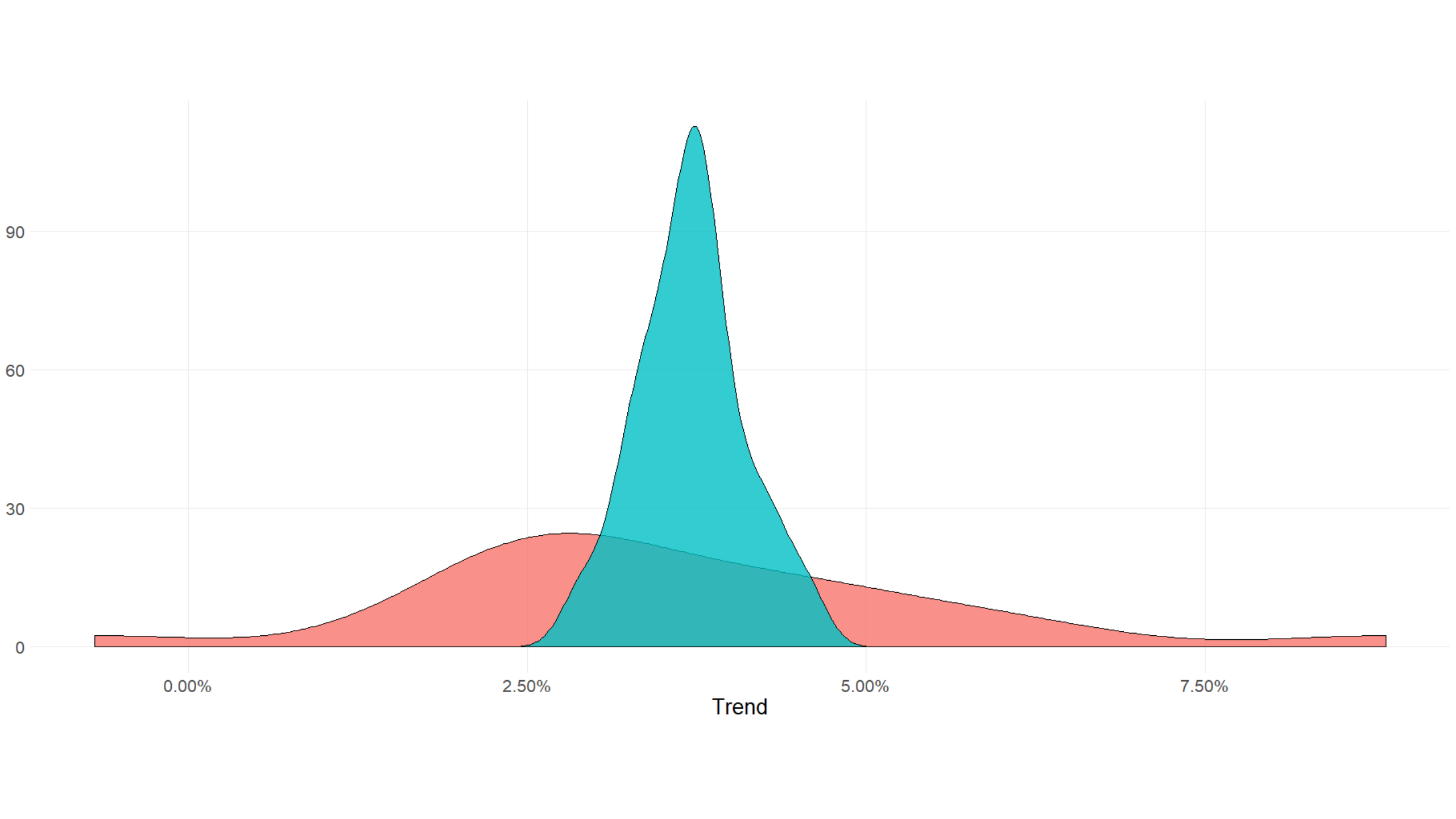
Yearly trend in claim index



Yearly trend in claim index

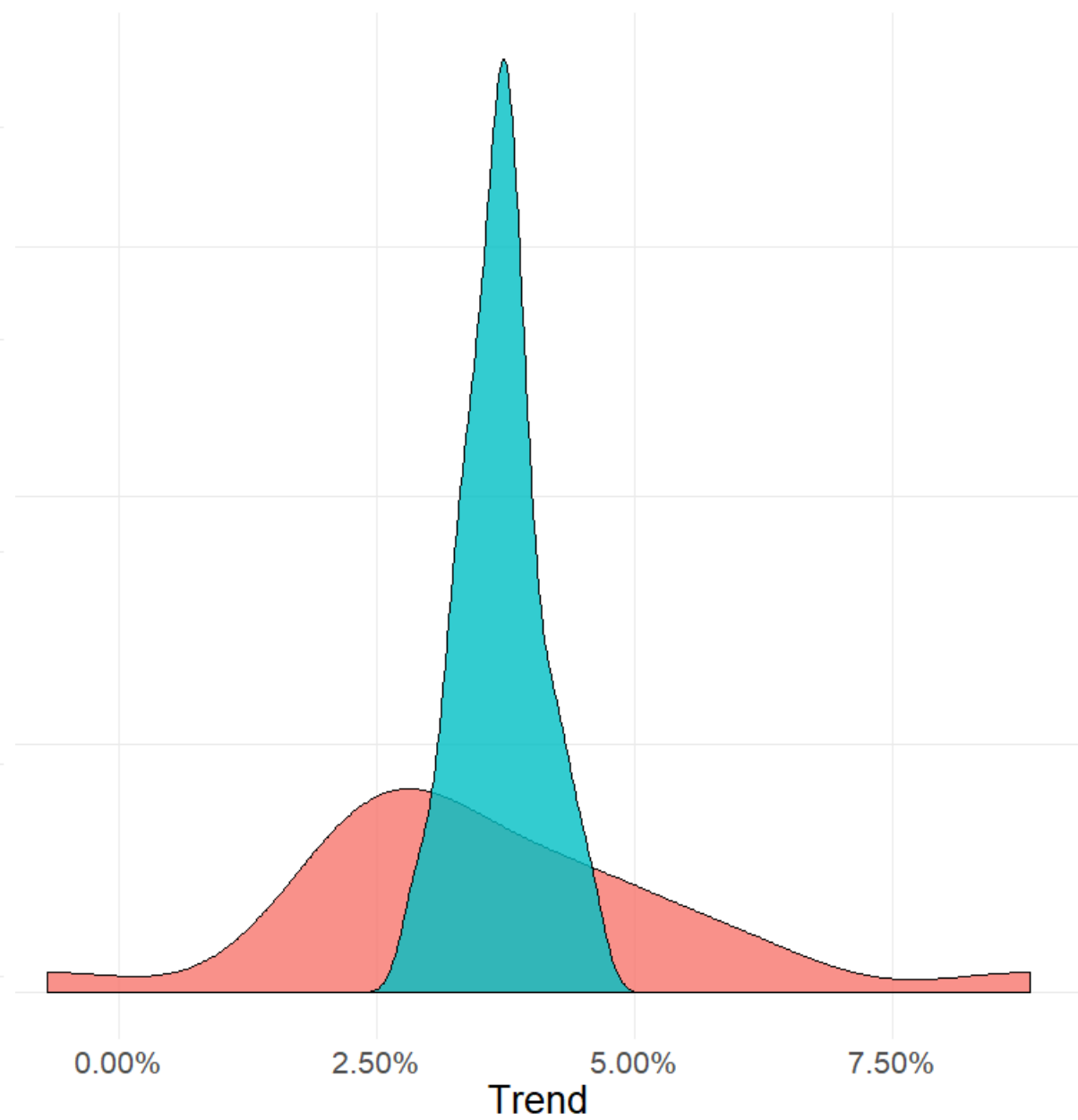
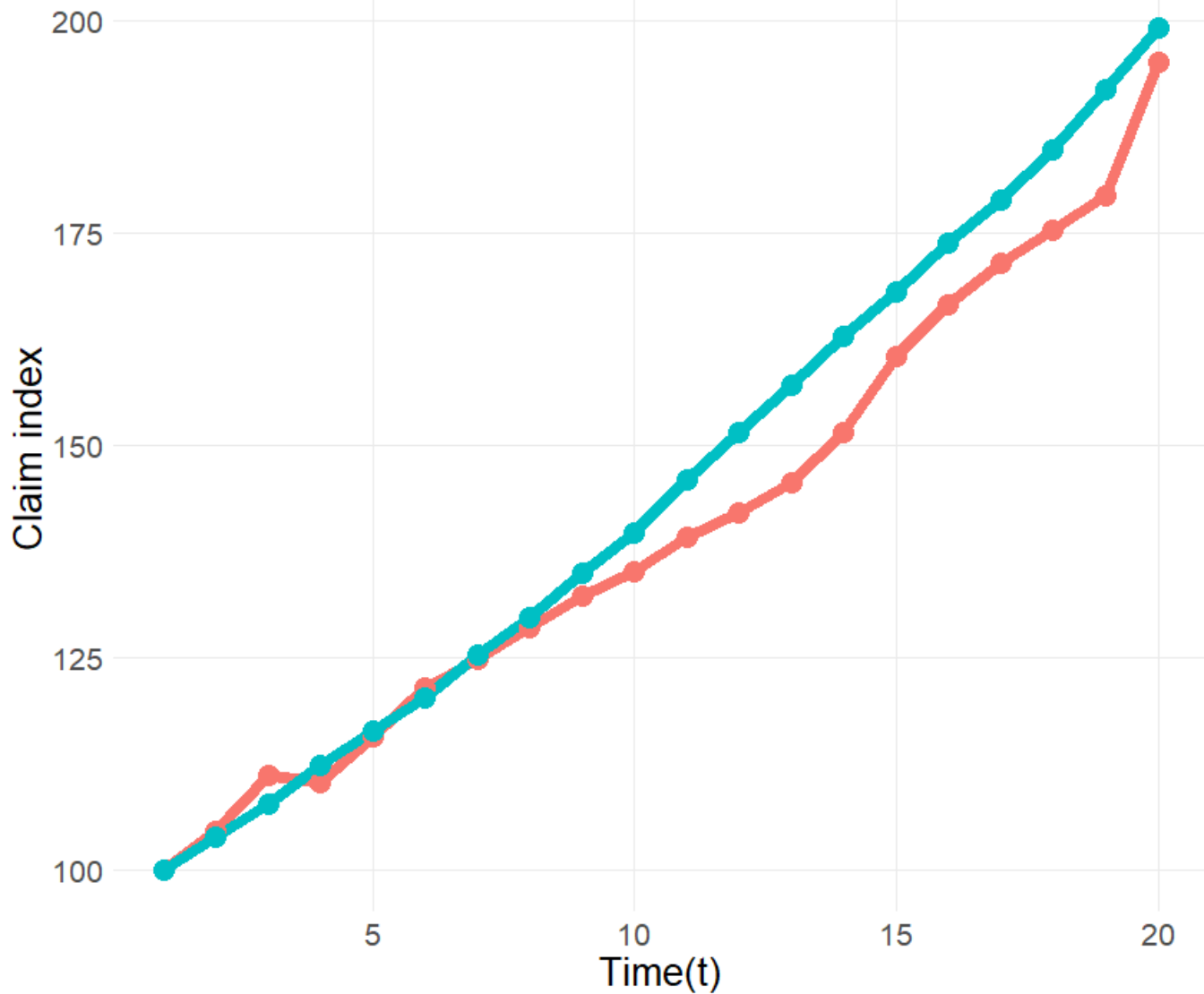
Values are shown in ascending order





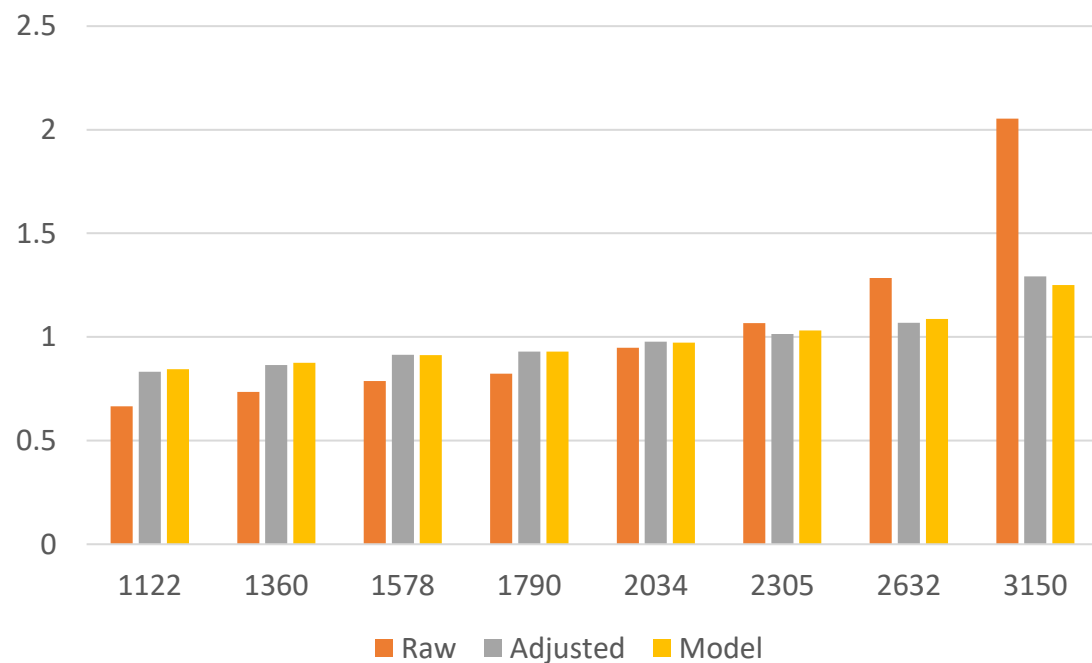
Relative mean claim size

Mean claim indexed to value of 100 at $t = 1$

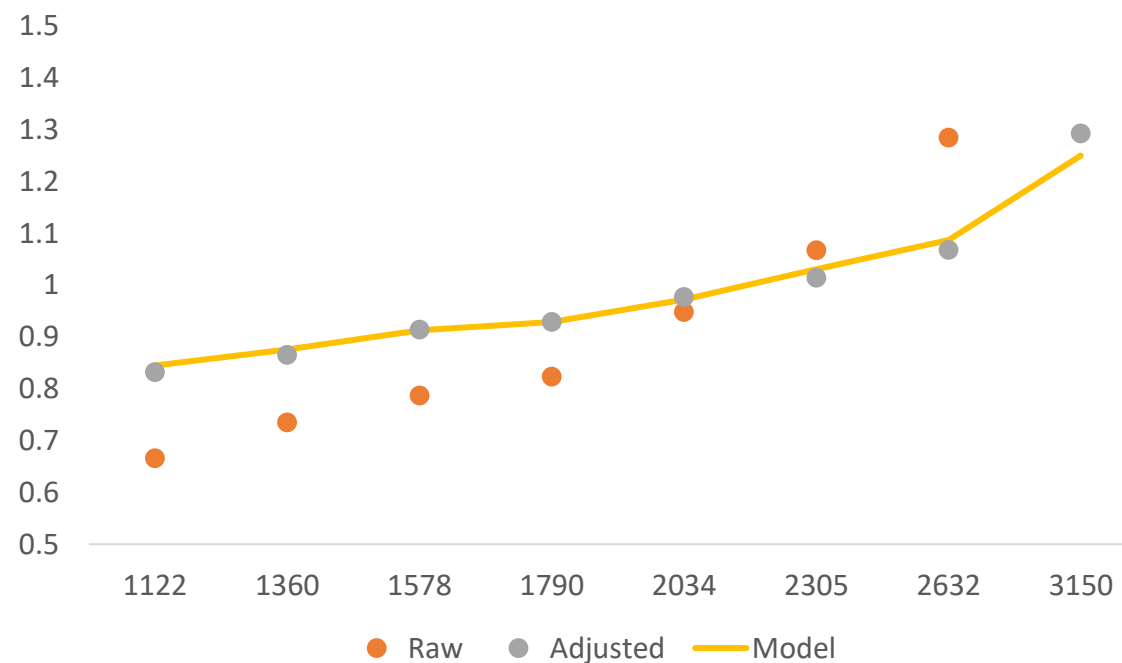


Rate That Visual!

Square Feet Living



Square Feet Living



Poll Question

- Which visual?
 - Answer: Left
 - Answer: Right

Value	Raw	Model
98001	0.522	0.779
98002	0.433	0.781
98003	0.553	0.779
98004	2.553	1.246
98005	1.473	1.161
98006	1.565	1.031
98007	1.109	1.095
98008	1.128	1.051
98010	0.787	0.821
98011	0.916	1.045
98014	0.764	0.928
98019	0.783	0.907
98022	0.579	0.85
98023	0.526	0.785
98024	1.085	0.909
98027	1.147	0.94
98028	0.849	0.893
98029	1.139	1.027
98030	0.557	0.798
98031	0.552	0.815
98032	0.486	0.812
98033	1.492	1.107
98034	0.937	1.049
98038	0.673	0.845
98039	4.264	1.218
98040	2.25	1.129
98042	0.568	0.807
98045	0.792	0.922
98052	1.193	1.046

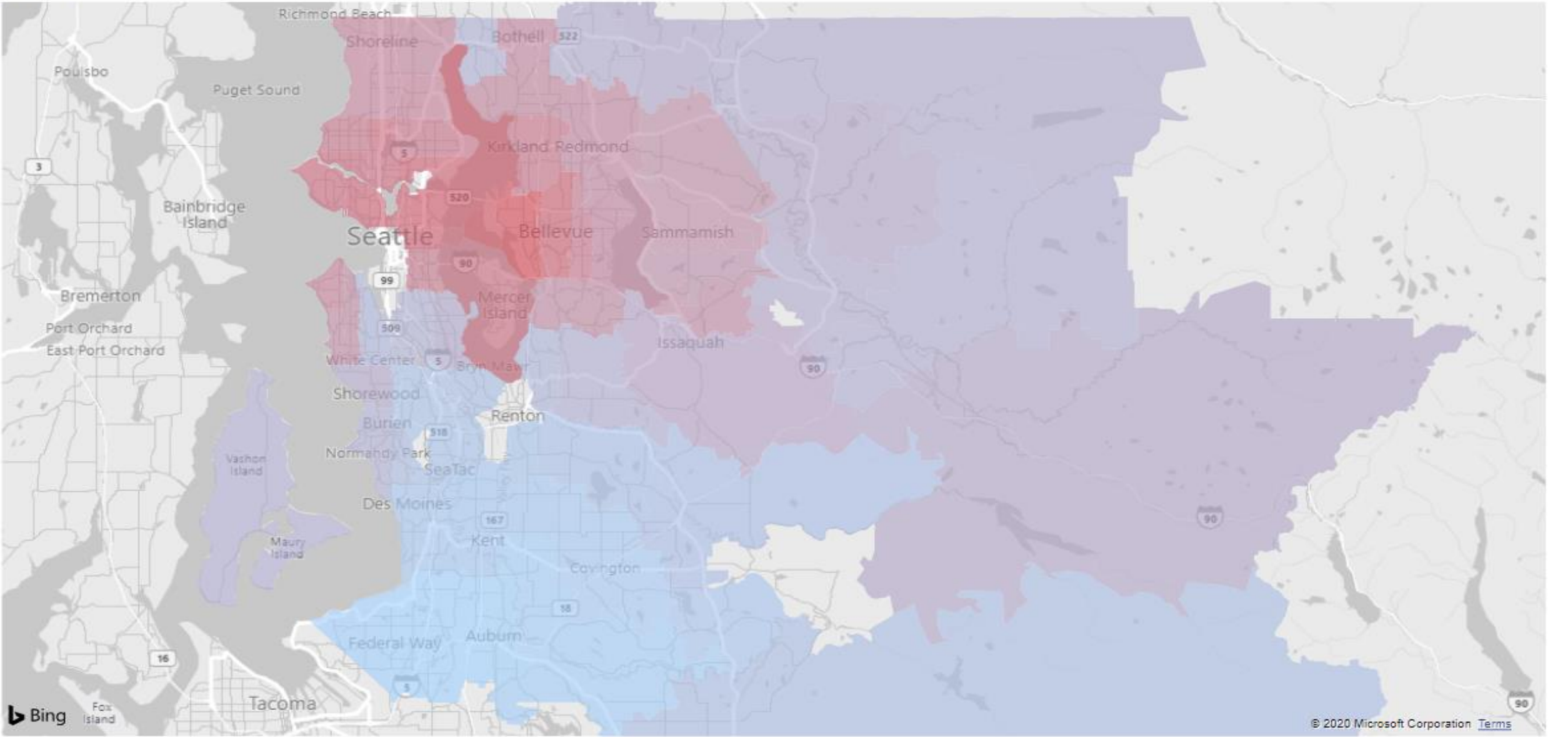
Value	Raw	Model
98001	0.522	0.779
98002	0.433	0.781
98003	0.553	0.779
98004	2.553	1.246
98005	1.473	1.161
98006	1.565	1.031
98007	1.109	1.095
98008	1.128	1.051
98010	0.787	0.821
98011	0.916	1.045
98014	0.764	0.928
98019	0.783	0.907
98022	0.579	0.85
98023	0.526	0.785
98024	1.085	0.909
98027	1.147	0.94
98028	0.849	0.893
98029	1.139	1.027
98030	0.557	0.798
98031	0.552	0.815
98032	0.486	0.812
98033	1.492	1.107
98034	0.937	1.049
98038	0.673	0.845
98039	4.264	1.218
98040	2.250	1.129
98042	0.568	0.807
98045	0.792	0.922
98052	1.193	1.046

Value	Raw	Model
98001	0.522	0.779
98002	0.433	0.781
98003	0.553	0.779
98004	2.553	1.246
98005	1.473	1.161
98006	1.565	1.031
98007	1.109	1.095
98008	1.128	1.051
98010	0.787	0.821
98011	0.916	1.045
98014	0.764	0.928
98019	0.783	0.907
98022	0.579	0.85
98023	0.526	0.785
98024	1.085	0.909
98027	1.147	0.94
98028	0.849	0.893
98029	1.139	1.027
98030	0.557	0.798
98031	0.552	0.815
98032	0.486	0.812
98033	1.492	1.107
98034	0.937	1.049
98038	0.673	0.845
98039	4.264	1.218
98040	2.25	1.129
98042	0.568	0.807
98045	0.792	0.922
98052	1.193	1.046

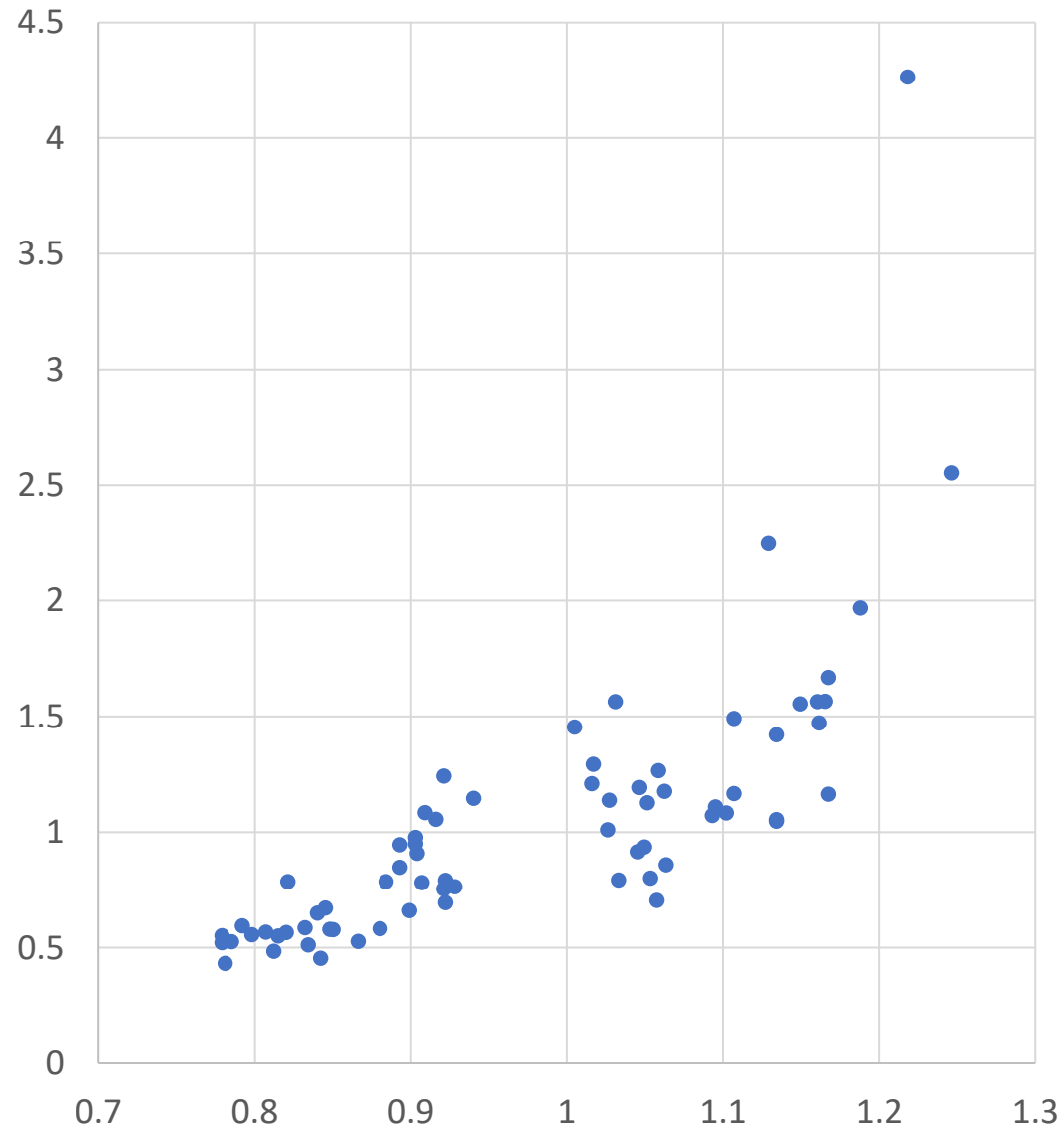
Poll Question

- Which visual?
 - Answer: Left
 - Answer: Right

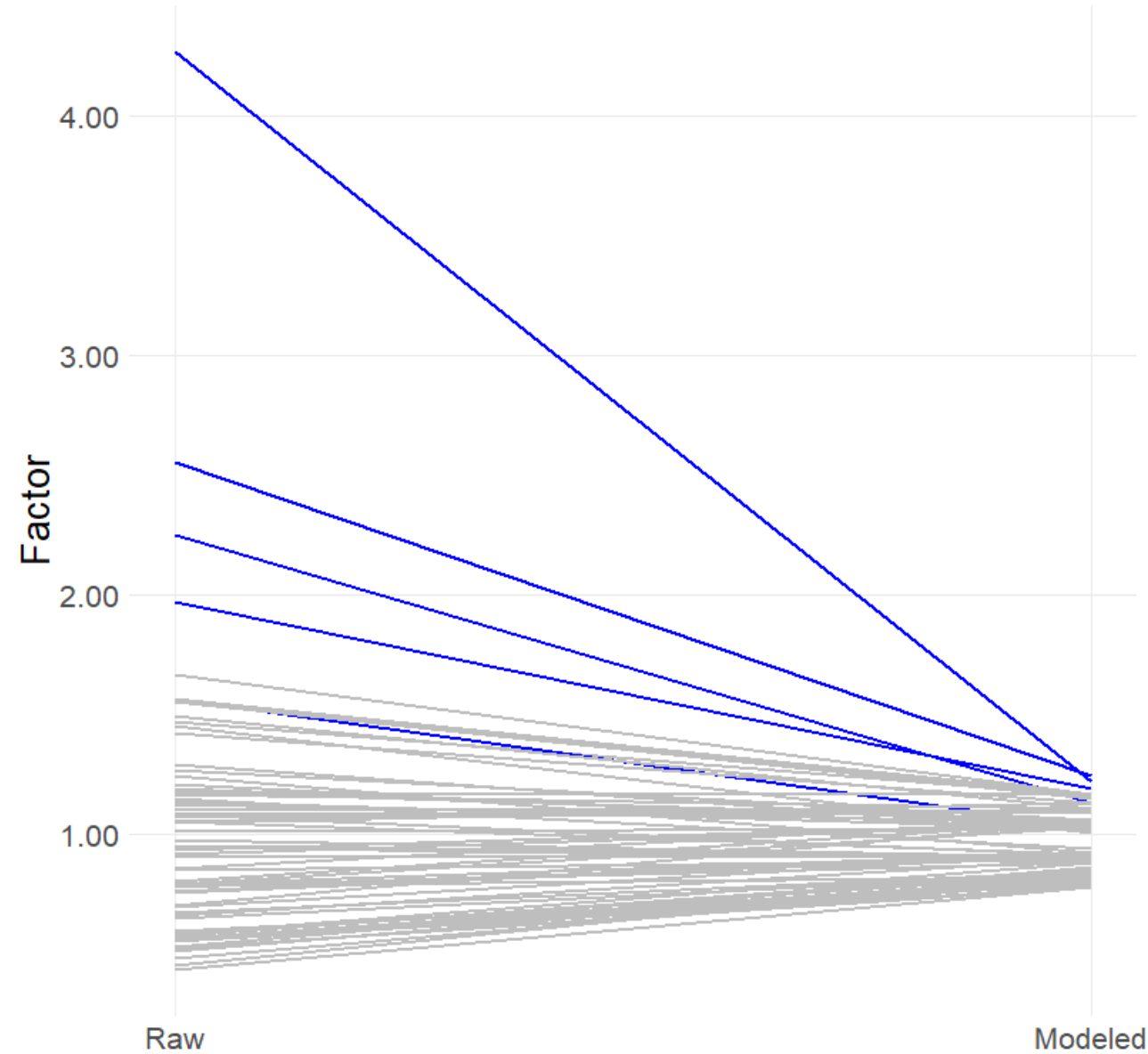
Zipcode



Raw (Y) vs Model (X) Factors - Zip Code



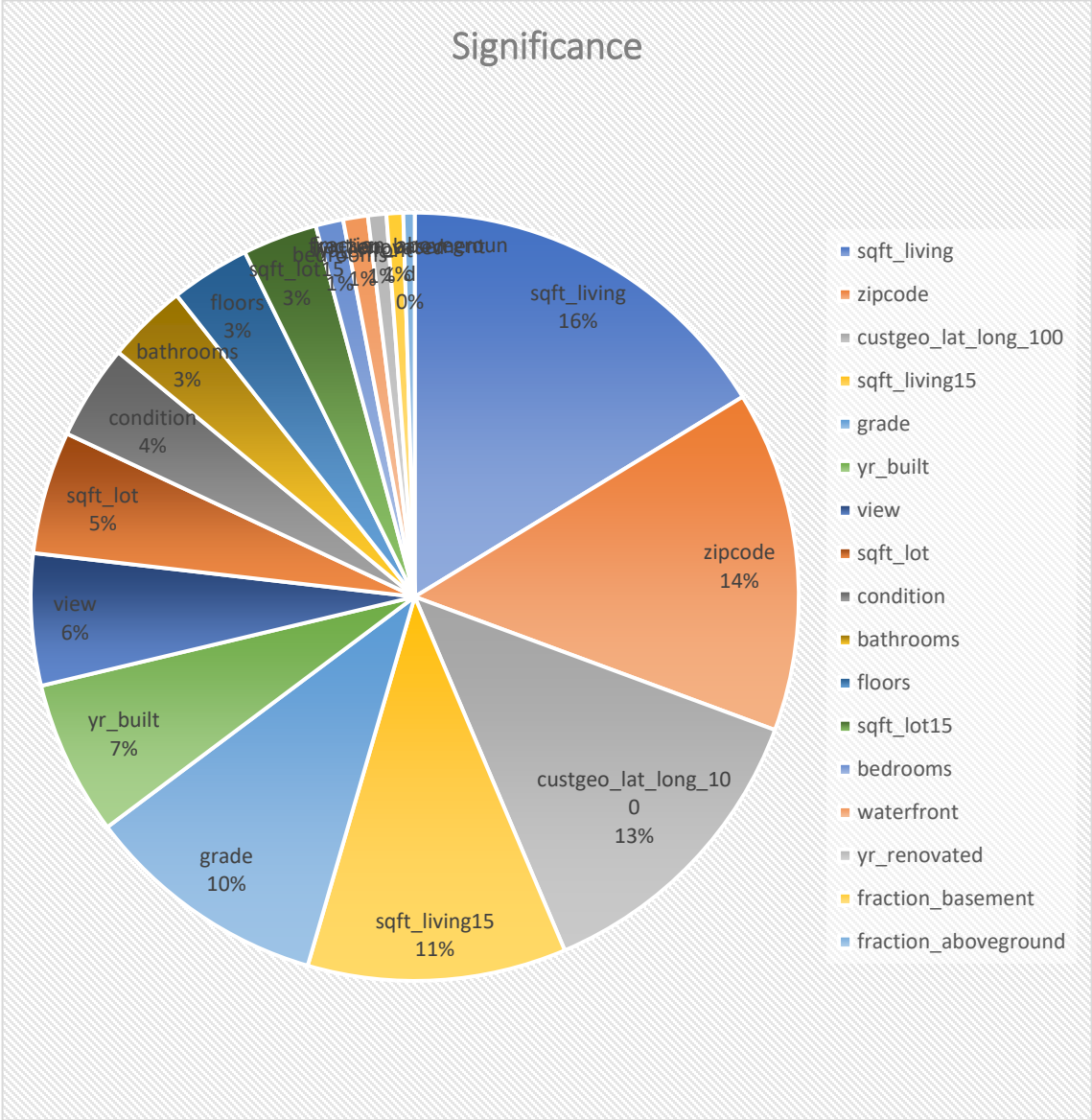
Raw vs. Model Factors - ZipCode



Poll Question

- Which visual?
 - Answer: Left
 - Answer: Right

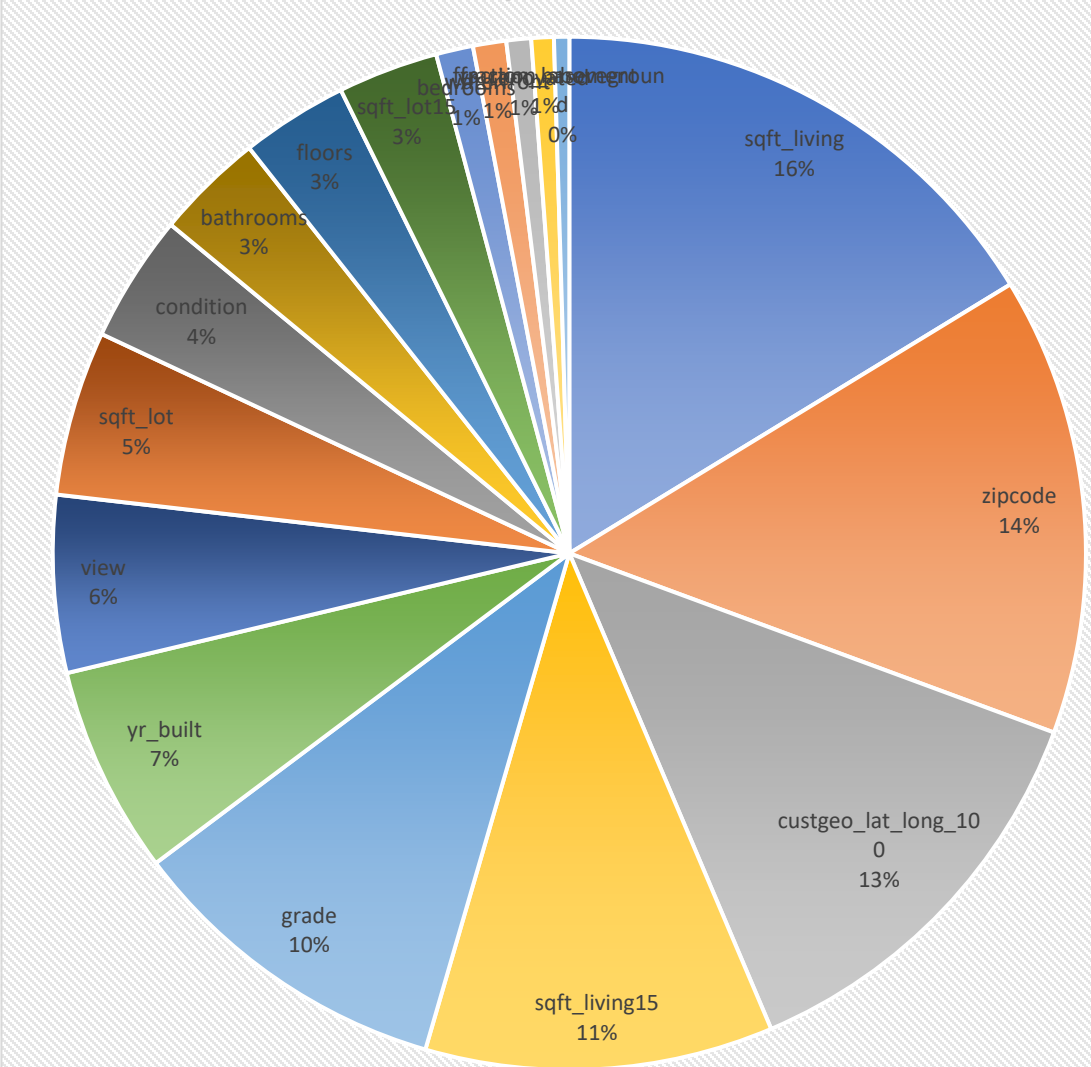
<u>category</u>	<u>Significance</u>
sqft_living	703,359,410
zipcode	619,875,488
custgeo_lat_long_100	560,447,484
sqft_living15	470,116,181
grade	443,707,937
yr_built	281,154,670
view	239,337,480
sqft_lot	222,928,937
condition	172,644,204
bathrooms	146,245,272
floors	144,536,108
sqft_lot15	135,665,143
bedrooms	50,115,418
waterfront	44,891,080
yr_renovated	33,403,891
fraction_basement	30,304,489
fraction_aboveground	20,562,045



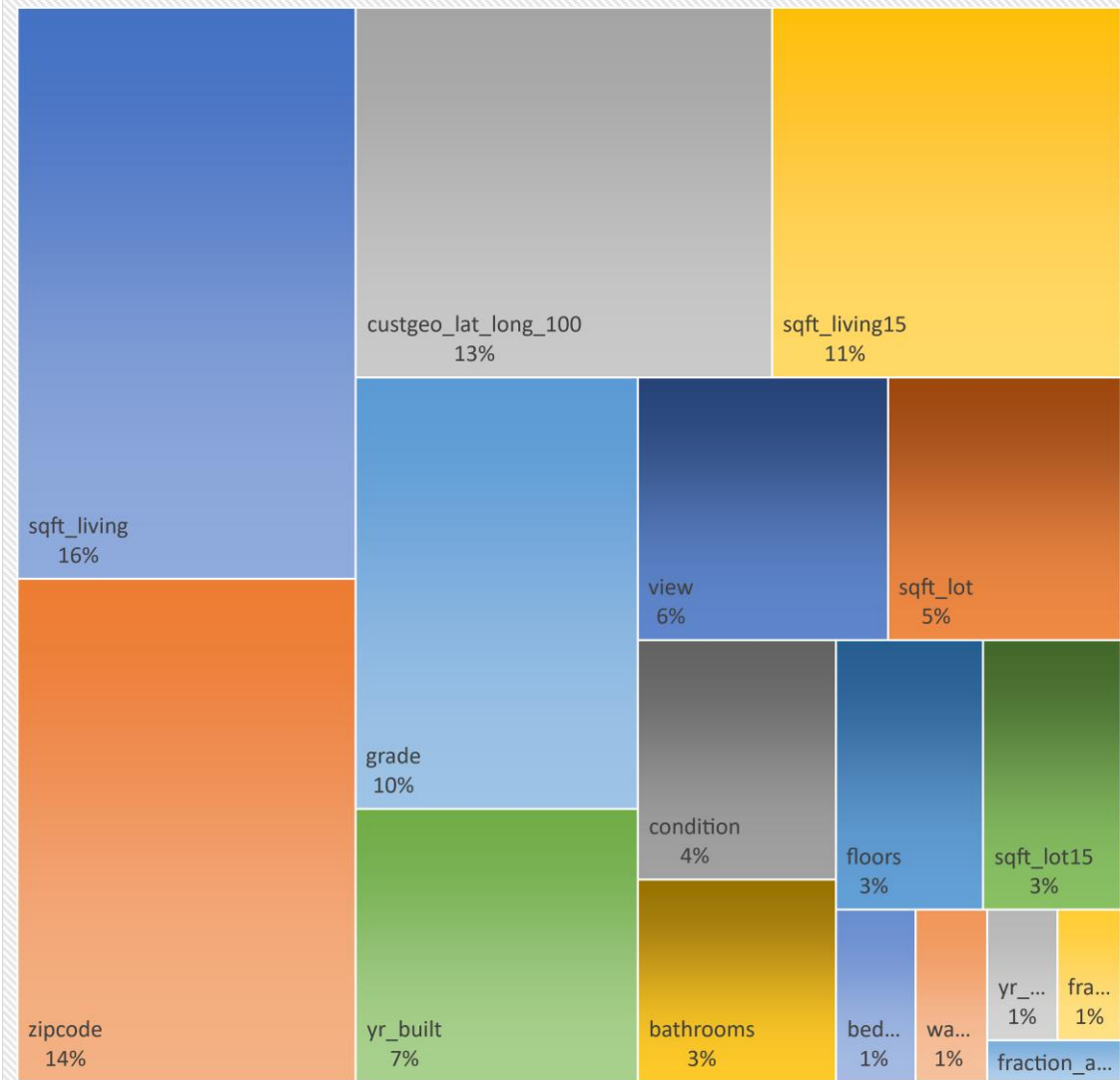
Poll Question

- Rate That Visual
 - Answer: Really?
 - Answer: Maybe
 - Answer: Tremendous!

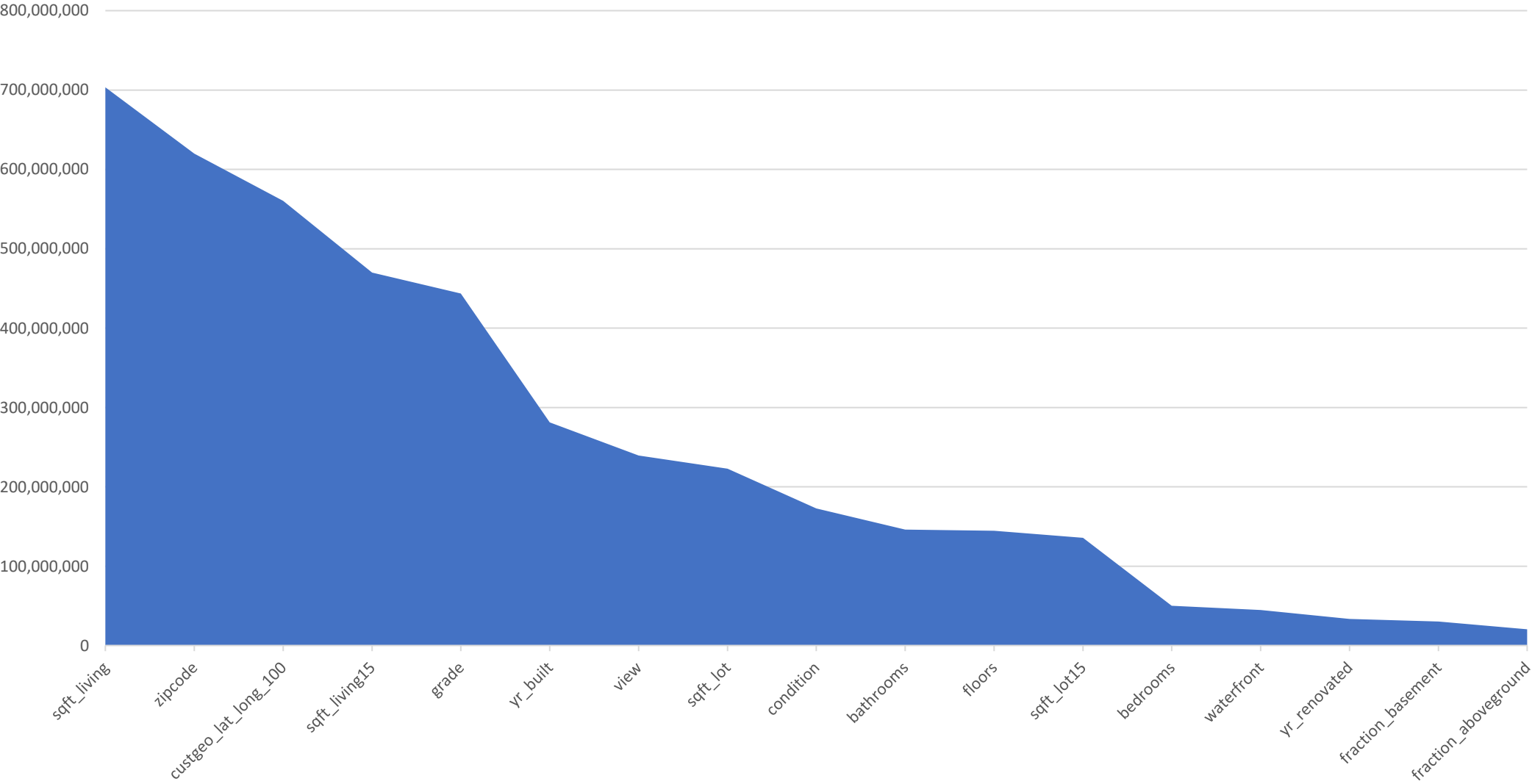
Significance



SIGNIFICANCE



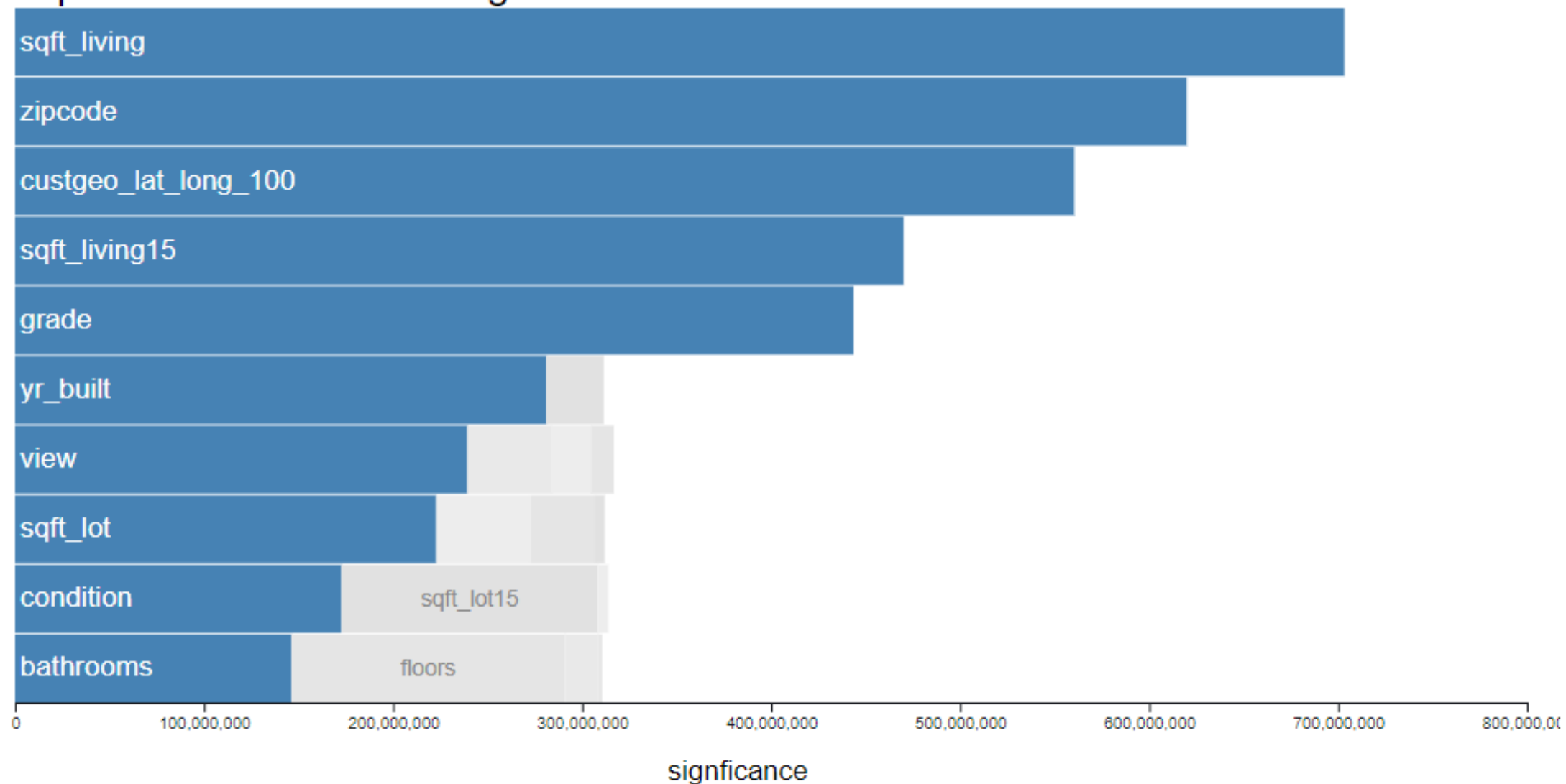
Top 5 variables are most significant



Poll Question

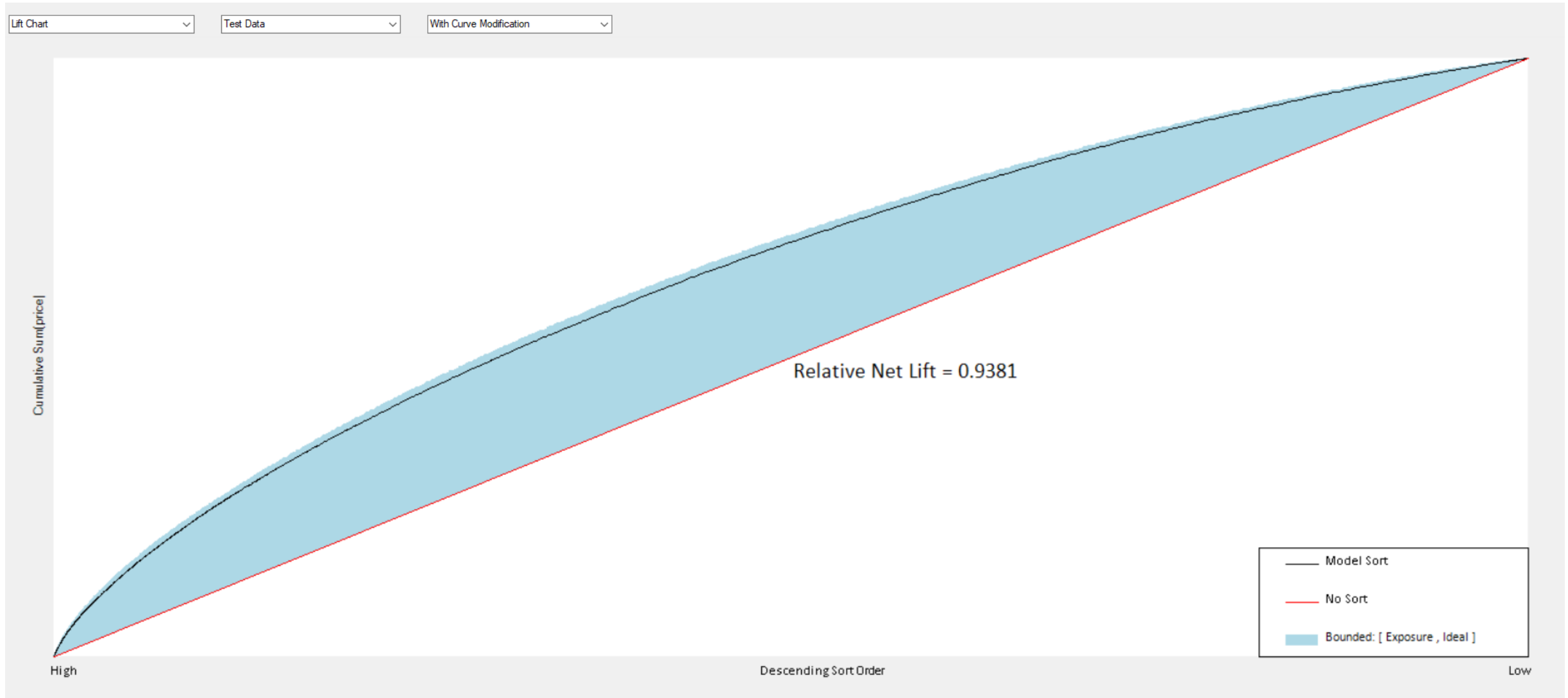
- Rate That Visual
 - Answer: Really?
 - Answer: Maybe
 - Answer: Tremendous!

Top 5 variables are most significant



Poll Question

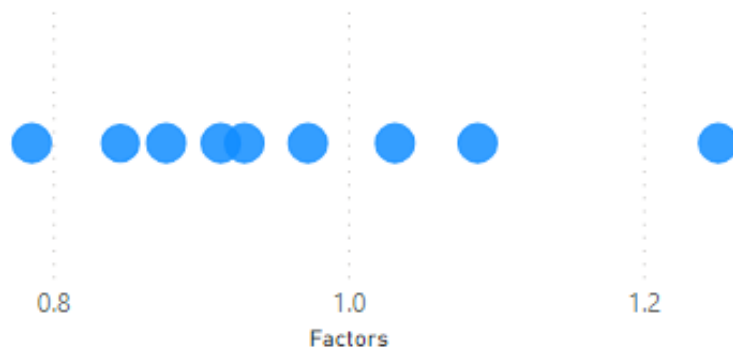
- Rate that visual?
 - Answer: Really?
 - Answer: Maybe
 - Answer: Tremendous!



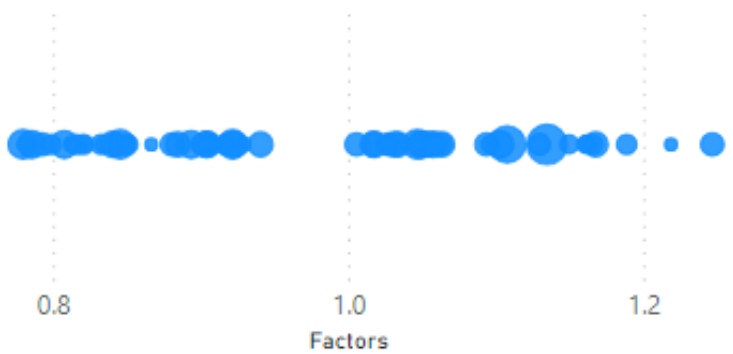
Poll Question

- Rate that visual?
 - Answer: Really?
 - Answer: Maybe
 - Answer: Tremendous!

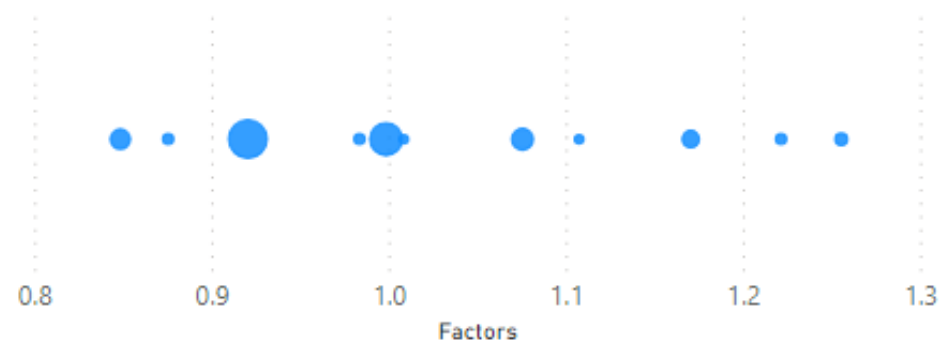
sqft_living



zipcode



grade



Poll Question

- Rate that visual?
 - Answer: Really?
 - Answer: Maybe
 - Answer: Tremendous!

Future Focused Working Parties

Visualization with respect to:

- Traditional Pricing
- Predictive Modeling
- Data Prep and Quality
- Business Leader Presentations

Resource List - Authors:

Knaflic, Cole Nusbaumer

Tufte, Edward

Few, Stephen

Munzner, Tamara

Cairo, Alberto

Yau, Nathan

Evergreen, Stephanie

Gregg, Xan