Statistical Graphics Support For Vega-Lite

Ayush Saraf

Problem

There was no support for ranged marks and therefore statistical graphics like error bars, box plots, or error bands in vega-lite.
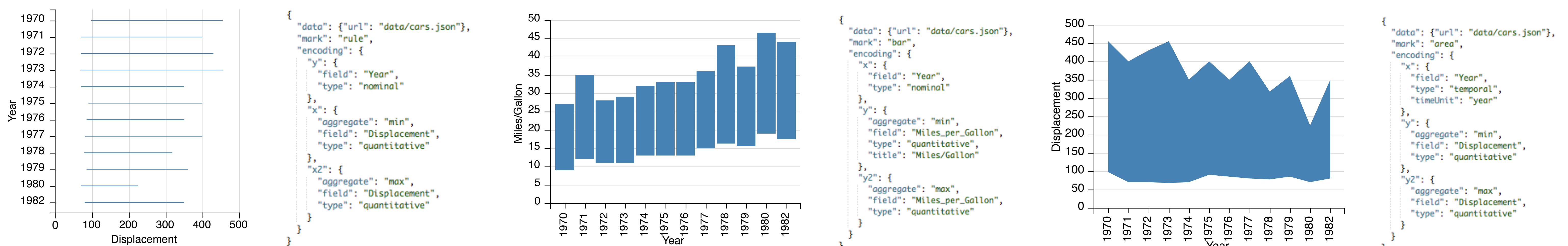
Motivation

Statistical graphics in data visualization would have tremendous applications in a data visualization toolkit like vega-lite. The most difficult aspect of this problem was design of the specification that represents these graphics for vega-lite.
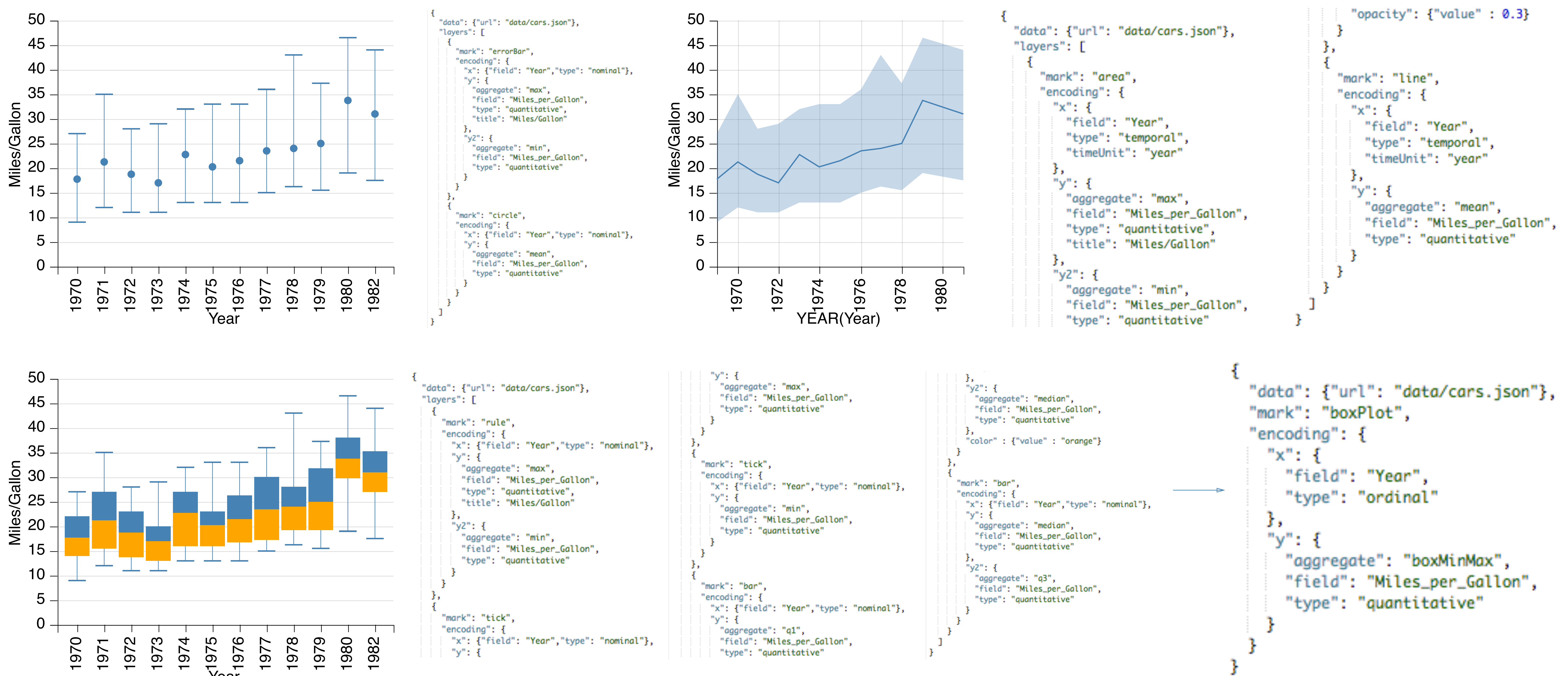
Approach

- Design a robust specification to support ranged rule, area and bar mark type.
- Implement and write test cases for the ranged marks.
- Use Ranged Marks with layering write various examples like error bars, error bands, and box plots.
- Design a concise specification for common statistical graphics like error bar and simple boxplot.
- Implement and write test cases for the composite marks which gets normalized to a layered specification.

Examples

New Primitive: Ranged Marks

New Composite: Statistical Graphics

Future Work

1. Currently you can only make a min max box and whiskers plot because of the limitations of the aggregation methods in Vega. Some future work may involve having supporting more variety of boxplots.
2. Support for violin plots.