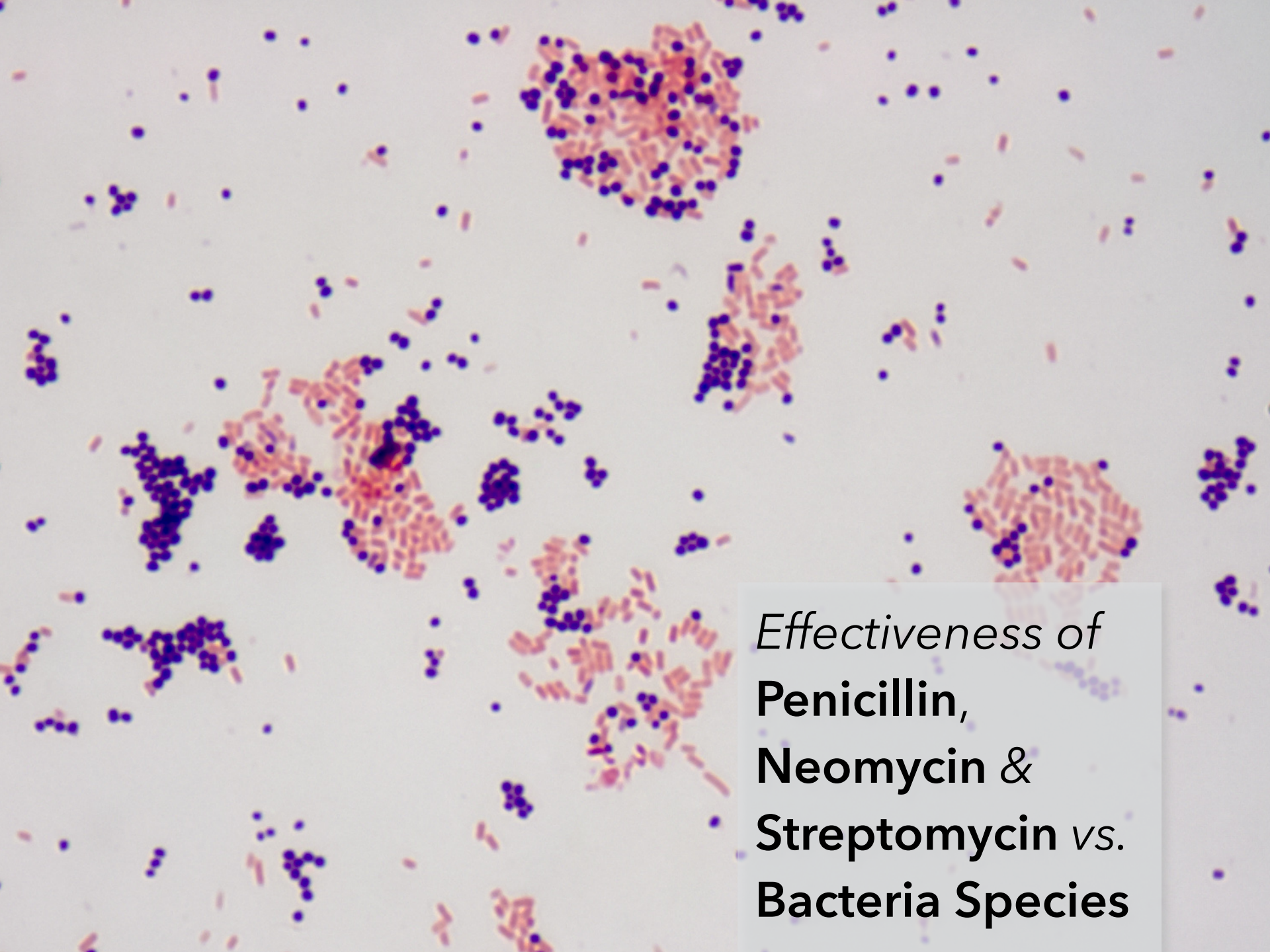
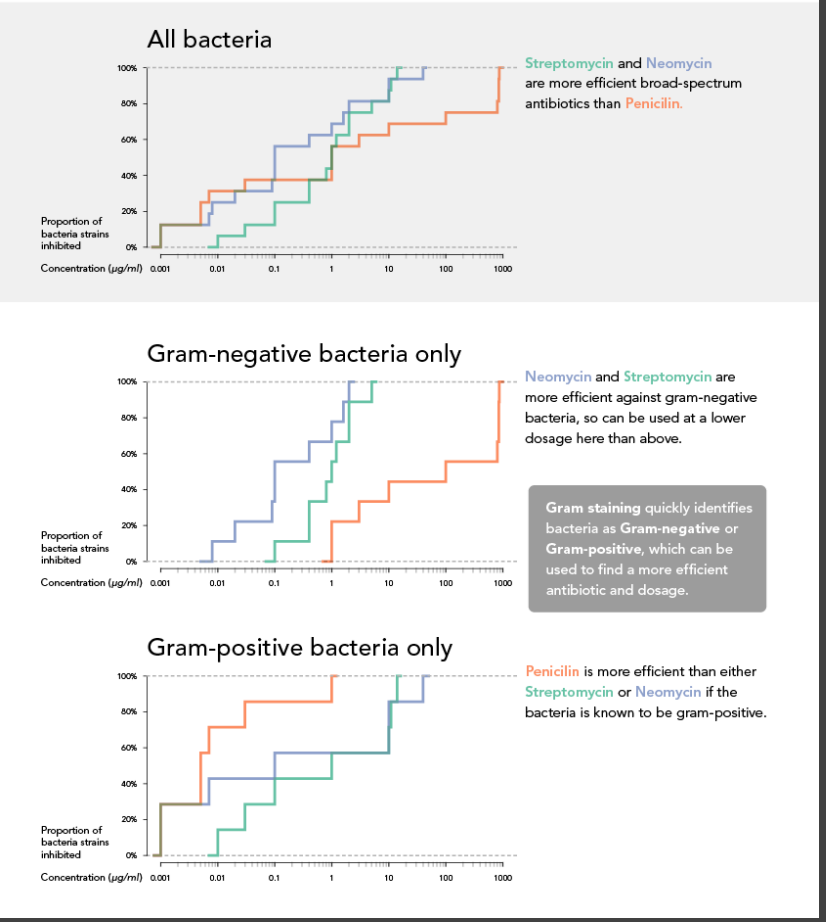
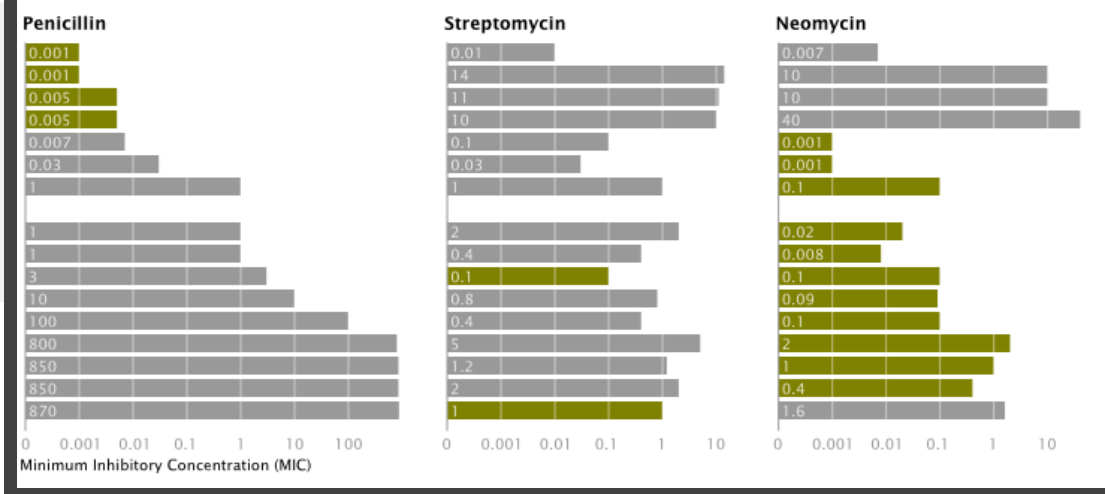
Visualization for DISCOVERY

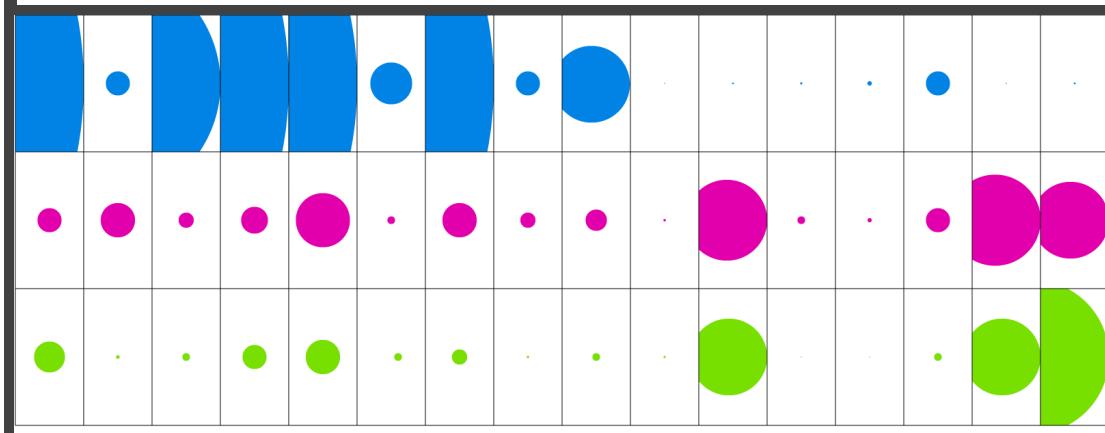
Jeffrey Heer @jeffrey_heer
Univ. of Washington + Trifacta

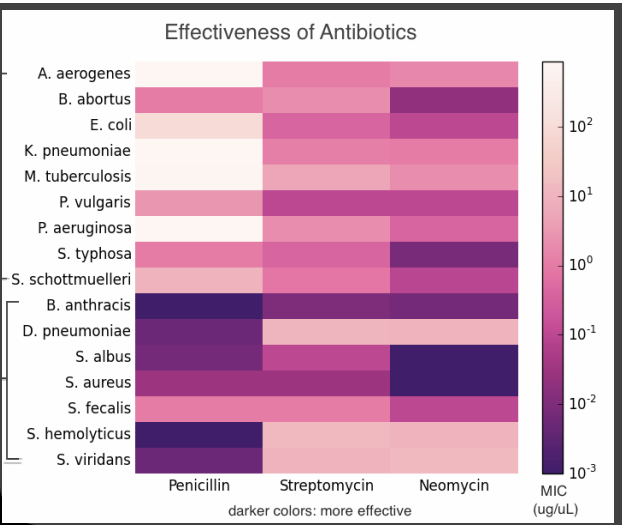

Effectiveness of
Penicillin,
Neomycin &
Streptomycin *vs.*
Bacteria Species

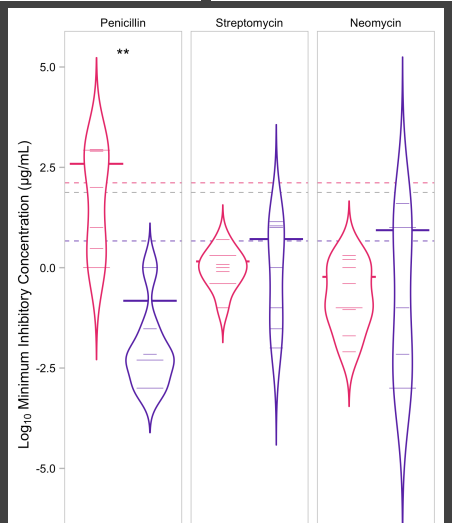
Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter aerogenes	870	1	1.6	negative
Brucella abortus	1	2	0.02	negative
Brucella anthracis	0.001	0.01	0.007	positive
Diplococcus pneumoniae	0.005	11	10	positive
Escherichia coli	100	0.4	0.1	negative
Klebsiella pneumoniae	850	1.2	1	negative
Mycobacterium tuberculosis	800	5	2	negative
Proteus vulgaris	3	0.1	0.1	negative
Pseudomonas aeruginosa	850	2	0.4	negative
Salmonella (Eberthella) typhosa	1	0.4	0.008	negative
Salmonella schottmuelleri	10	0.8	0.09	negative
Staphylococcus albus	0.007	0.1	0.001	positive
Staphylococcus aureus	0.03	0.03	0.001	positive
Streptococcus fecalis	1	1	0.1	positive
Streptococcus hemolyticus	0.001	14	10	positive
Streptococcus viridans	0.005	10	40	positive

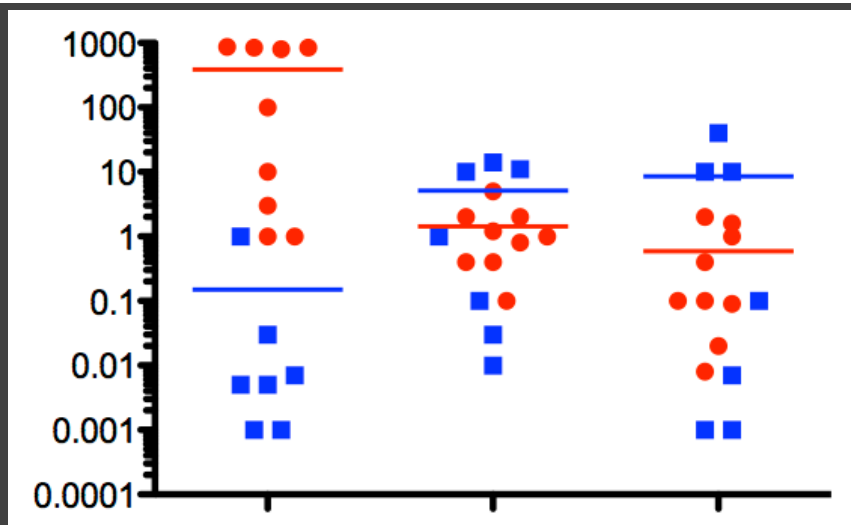

All bacteria
Streptomycin and Neomycin are more efficient broad-spectrum antibiotics than Penicillin.
Proportion of bacteria strains inhibited
Concentration (μg/ml)
Gram-negative bacteria only
Neomycin and Streptomycin are more efficient against gram-negative bacteria, so can be used at a lower dosage here than above.
Gram staining quickly identifies bacteria as Gram-negative or Gram-positive, which can be used to find a more efficient antibiotic and dosage.
Gram-positive bacteria only
Penicillin is more efficient than either Streptomycin or Neomycin if the bacteria is known to be gram-positive.

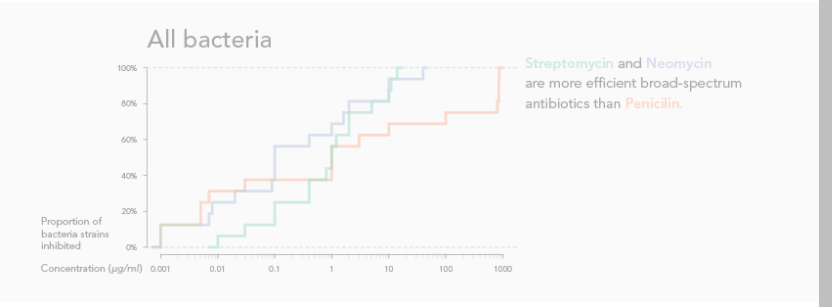

Penicillin
Streptomycin
Neomycin
Minimum Inhibitory Concentration (MIC)

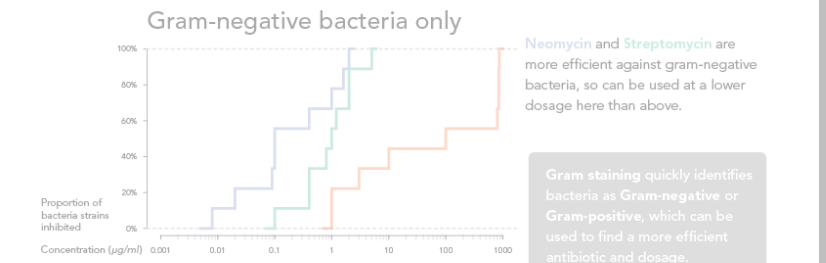


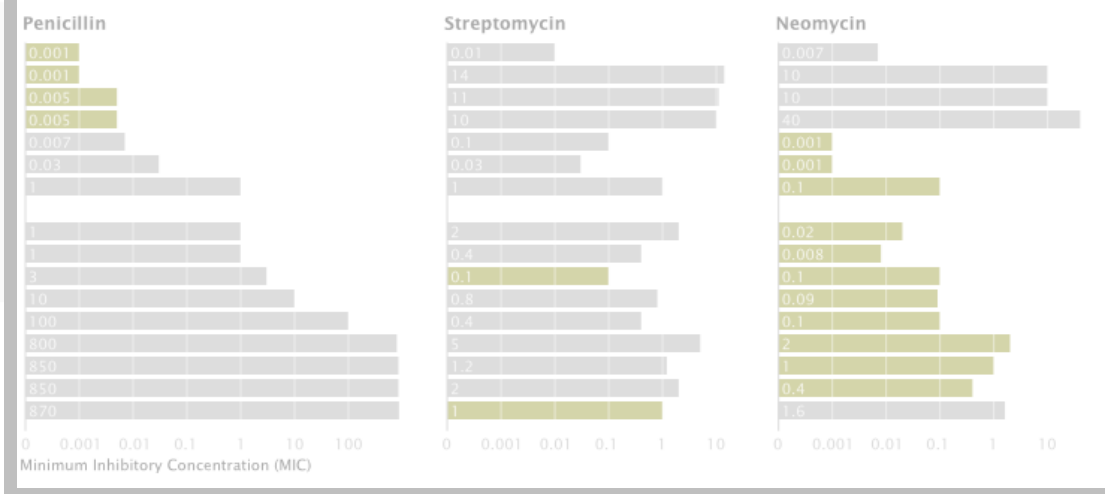

Effectiveness of Antibiotics
A. aerogenes
B. abortus
E. coli
K. pneumoniae
M. tuberculosis
P. vulgaris
P. aeruginosa
S. typhosa
S. schottmuelleri
B. anthracis
D. pneumoniae
S. albus
S. aureus
S. fecalis
S. hemolyticus
S. viridans
Penicillin
Streptomycin
Neomycin
darker colors: more effective
MIC (ug/uL)


Penicillin
Streptomycin
Neomycin
**
Log10 Minimum Inhibitory Concentration (μg/mL)

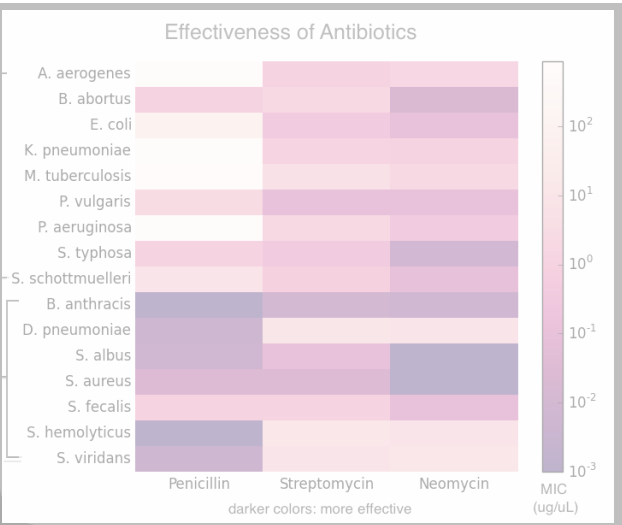

1000
100
10
1
0.1
0.01
0.001
0.0001

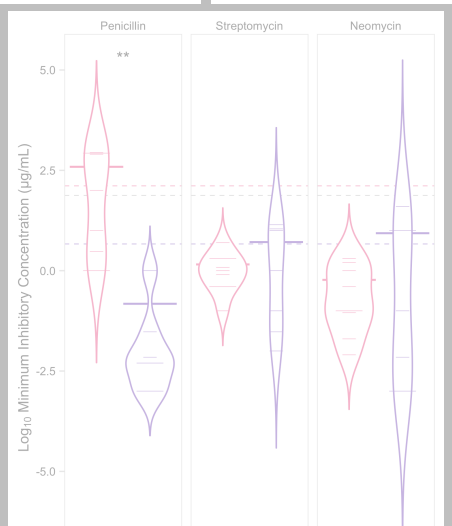


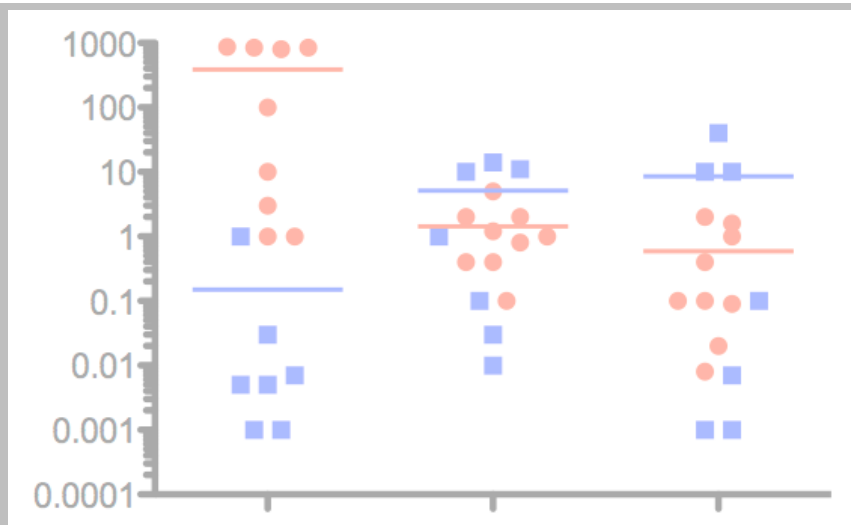




Which antibiotic should one use?

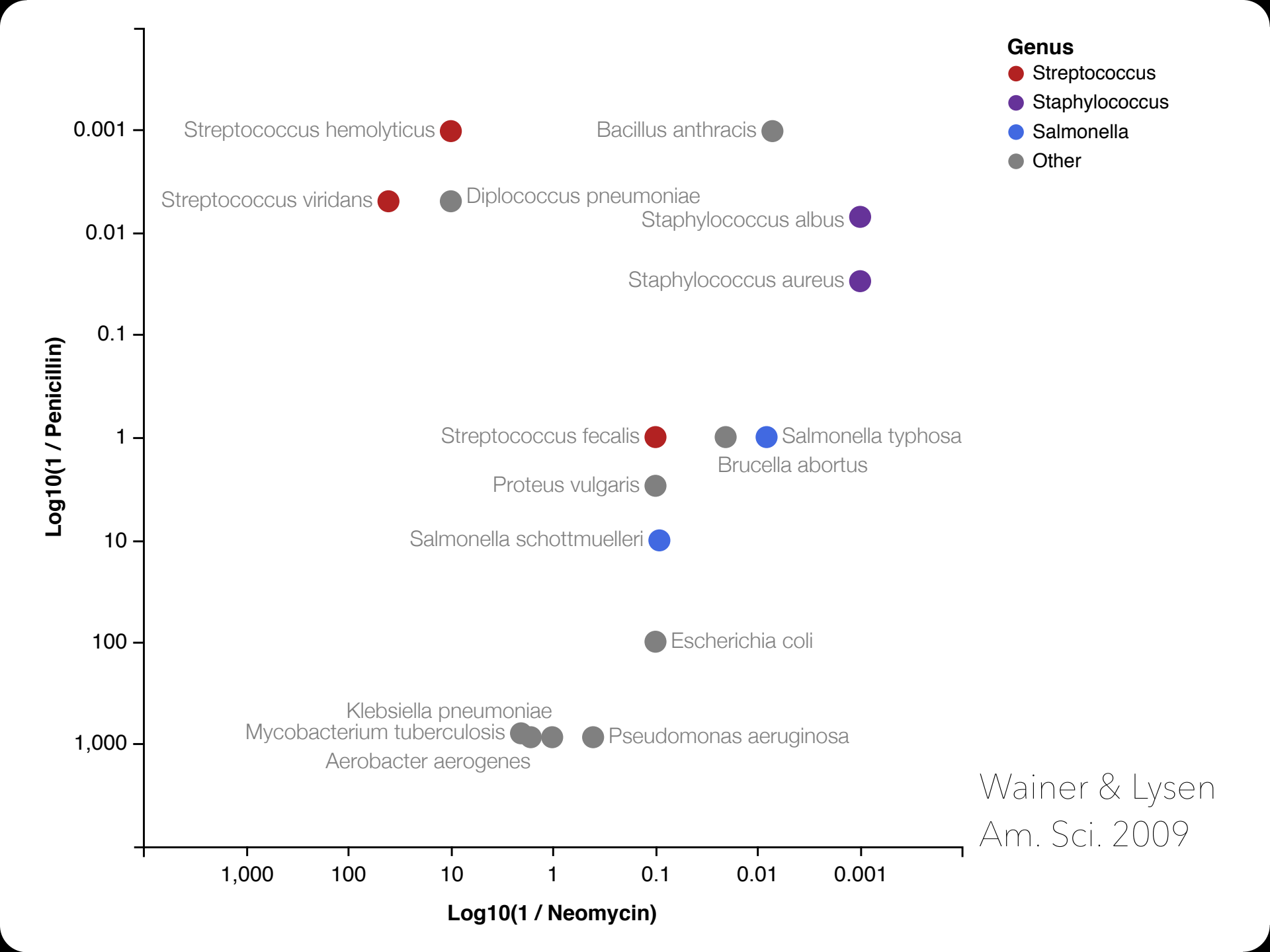


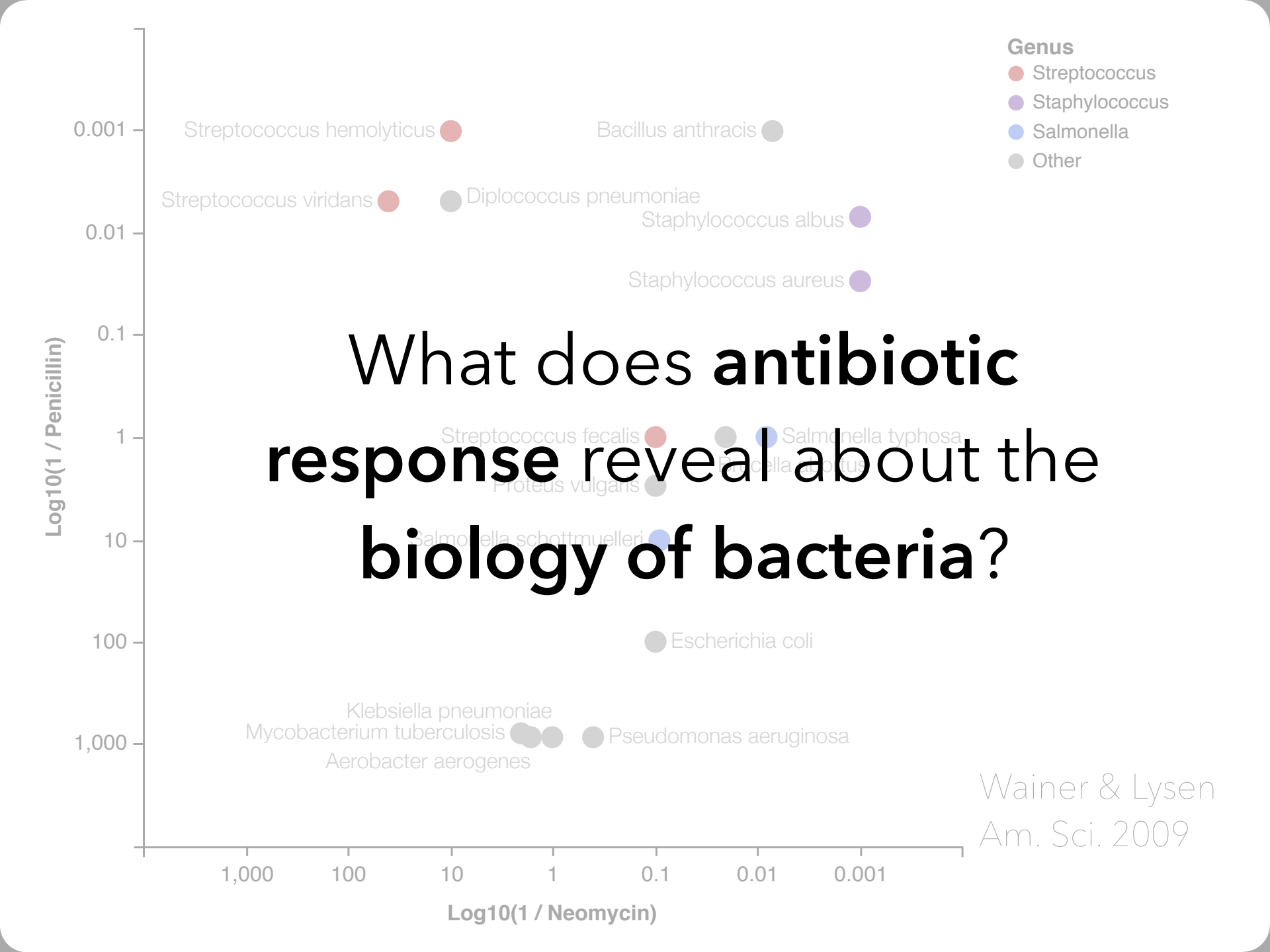




Show **data variation**,
not **design variation.**

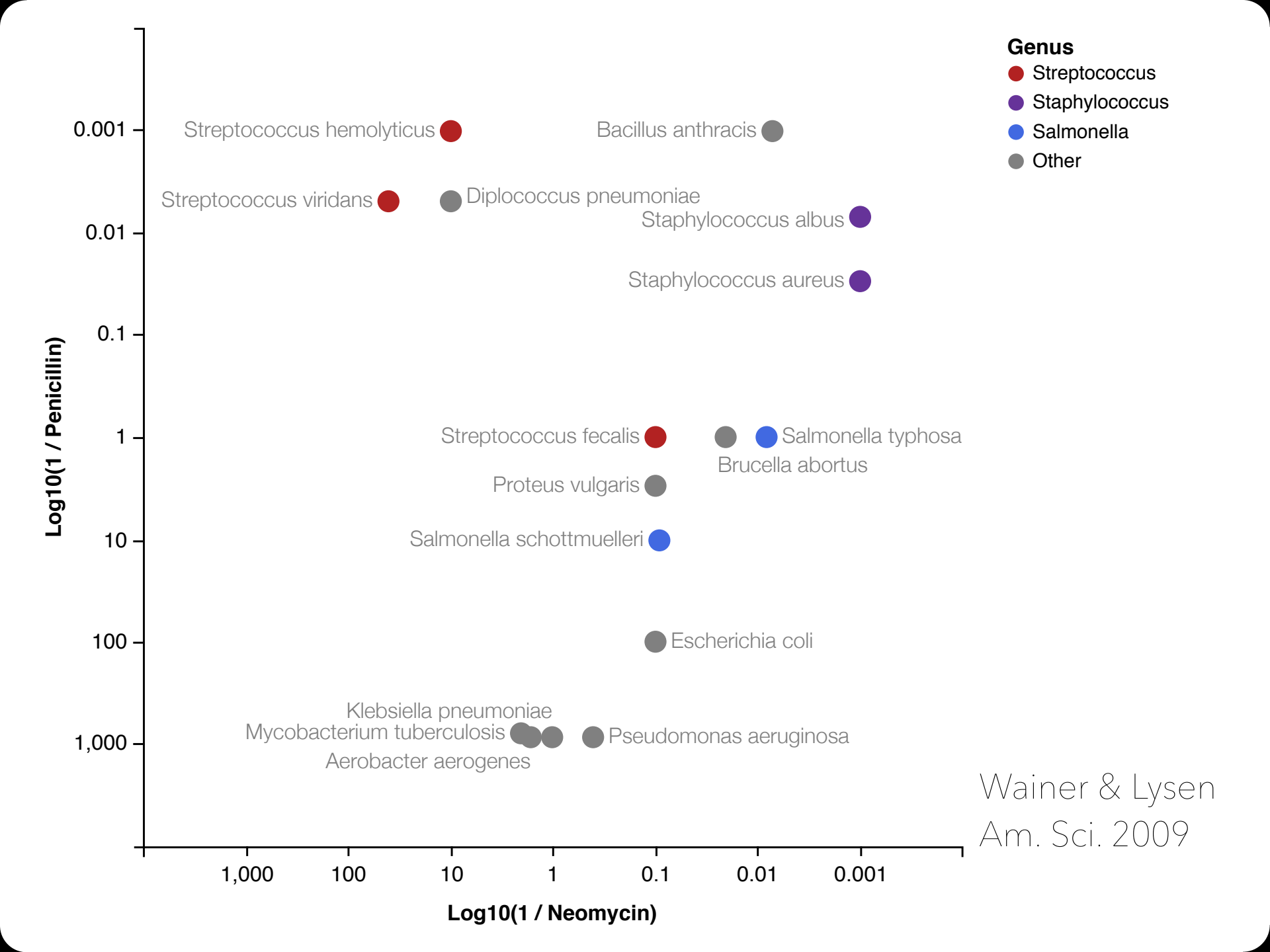
- Edward Tufte

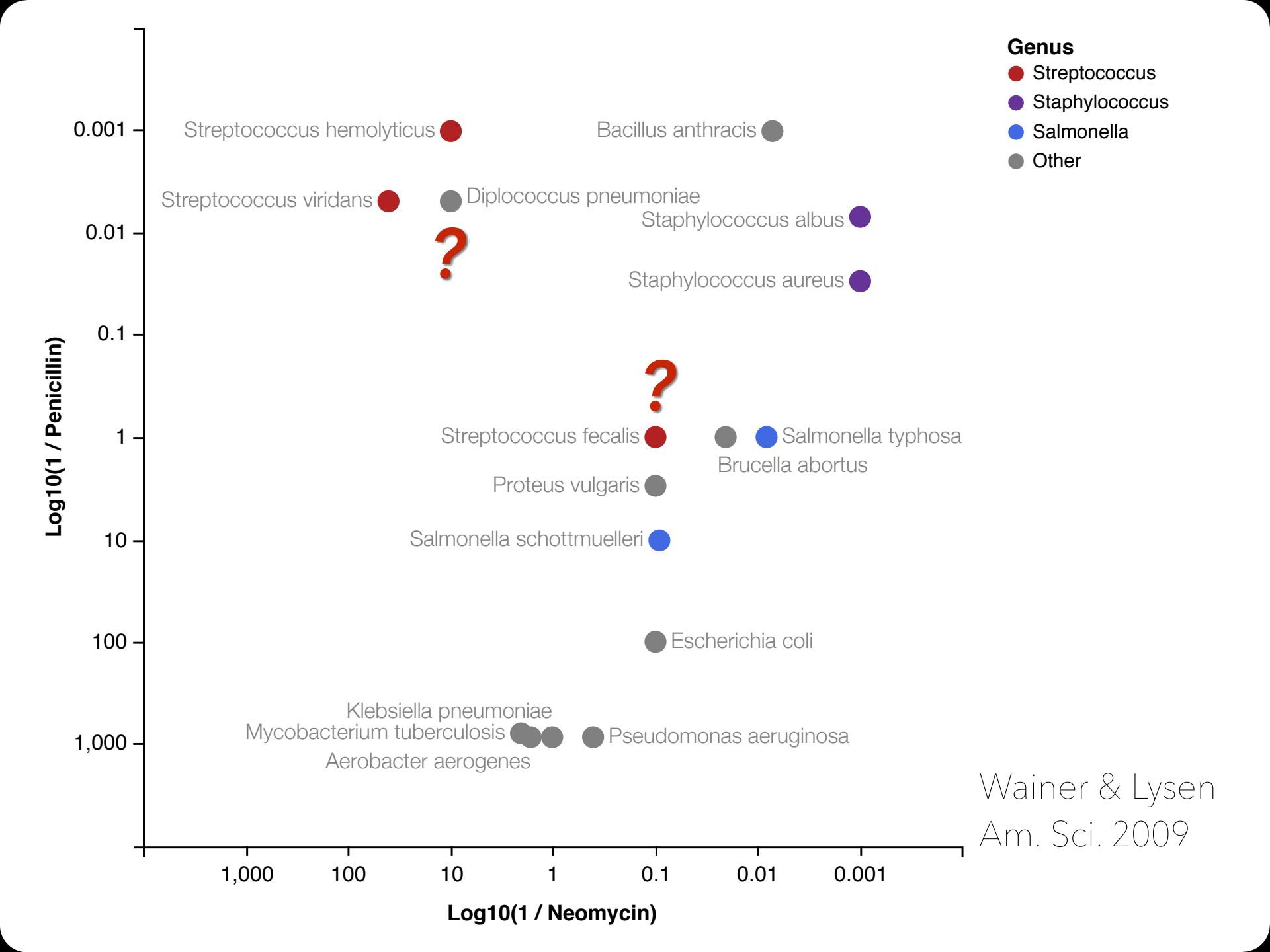

Genus
Streptococcus
Staphylococcus
Salmonella
Other
Streptococcus hemolyticus
Bacillus anthracis
Streptococcus viridans
Diplococcus pneumoniae
Staphylococcus albus
Staphylococcus aureus
Streptococcus fecalis
Salmonella typhosa
Brucella abortus
Proteus vulgaris
Salmonella schottmuelleri
Escherichia coli
Klebsiella pneumoniae
Mycobacterium tuberculosis
Aerobacter aerogenes
Pseudomonas aeruginosa
0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
1,000
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)
Wainer & Lysen
Am. Sci. 2009

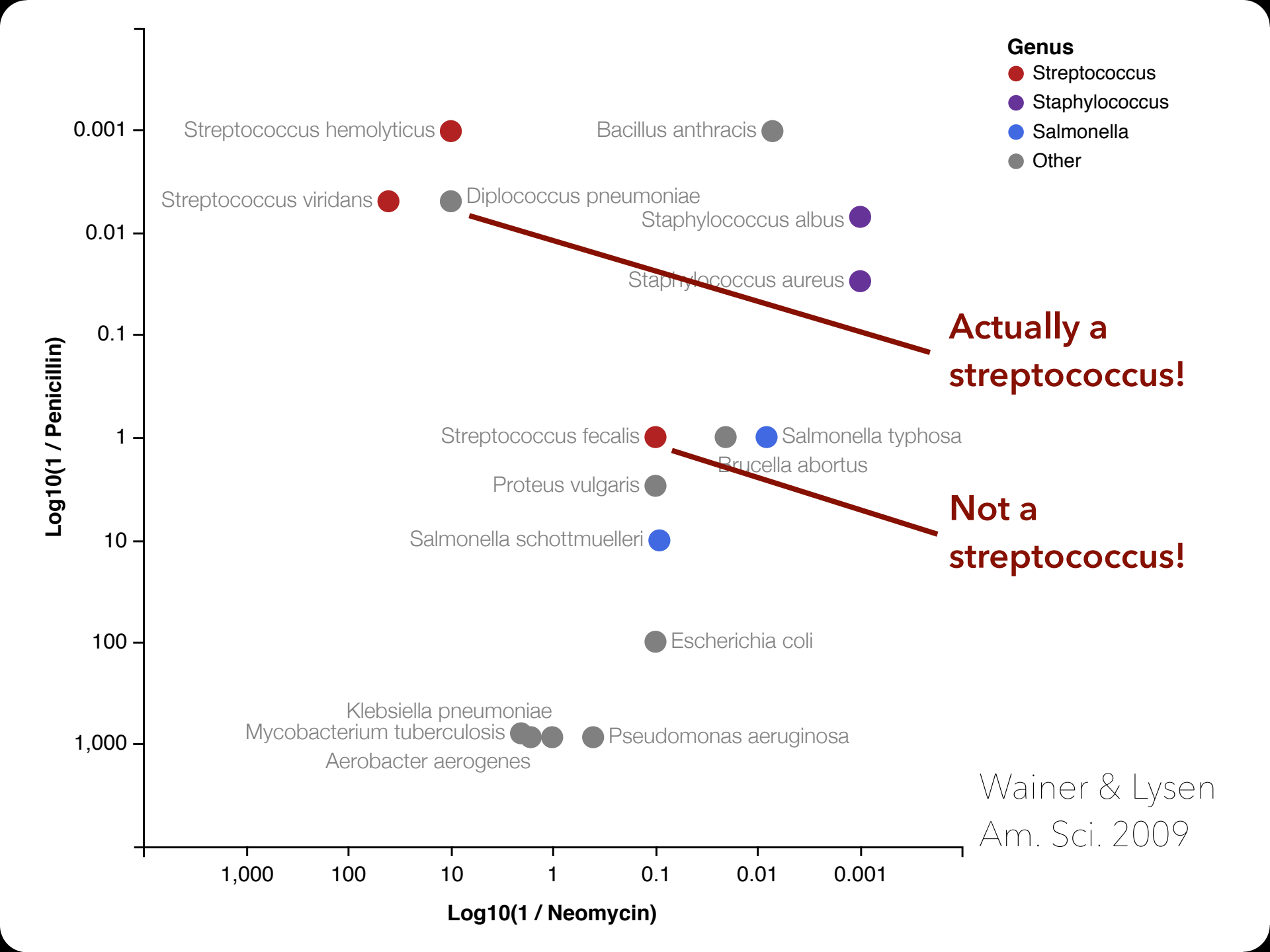


What does **antibiotic response** reveal about the **biology of bacteria**?

Wainer & Lysen
Am. Sci. 2009


Genus
Streptococcus
Staphylococcus
Salmonella
Other
Streptococcus hemolyticus
Bacillus anthracis
Streptococcus viridans
Diplococcus pneumoniae
Staphylococcus albus
Staphylococcus aureus
Streptococcus fecalis
Salmonella typhosa
Brucella abortus
Proteus vulgaris
Salmonella schottmuelleri
Escherichia coli
Klebsiella pneumoniae
Mycobacterium tuberculosis
Aerobacter aerogenes
Pseudomonas aeruginosa
0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
1,000
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)
Wainer & Lysen
Am. Sci. 2009

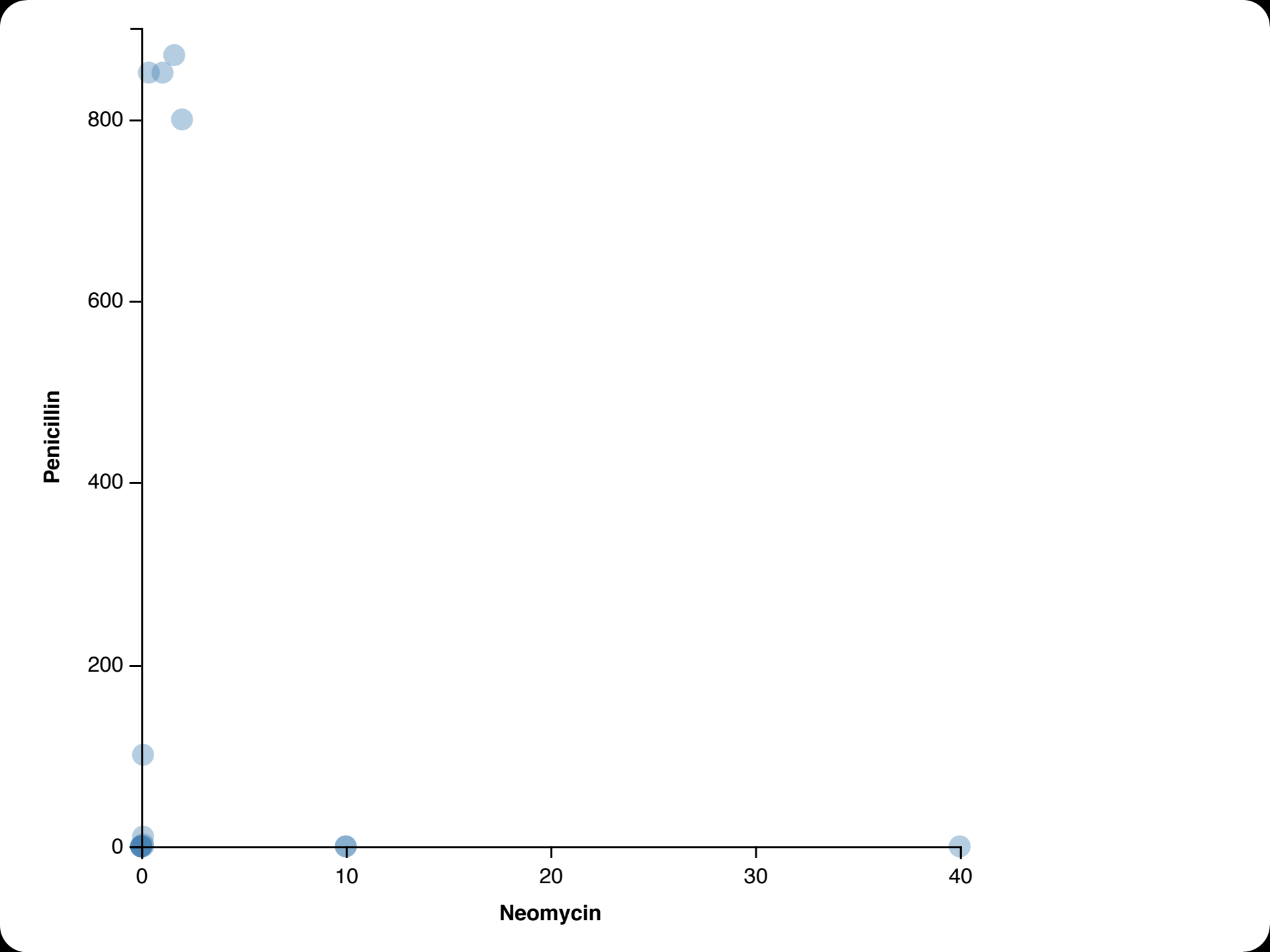

Genus
Streptococcus
Staphylococcus
Salmonella
Other
Streptococcus hemolyticus
Bacillus anthracis
Streptococcus viridans
Diplococcus pneumoniae
Staphylococcus albus
?
Staphylococcus aureus
?
Streptococcus fecalis
Salmonella typhosa
Brucella abortus
Proteus vulgaris
Salmonella schottmuelleri
Escherichia coli
Klebsiella pneumoniae
Mycobacterium tuberculosis
Aerobacter aerogenes
Pseudomonas aeruginosa
0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
1,000
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)
Wainer & Lysen
Am. Sci. 2009

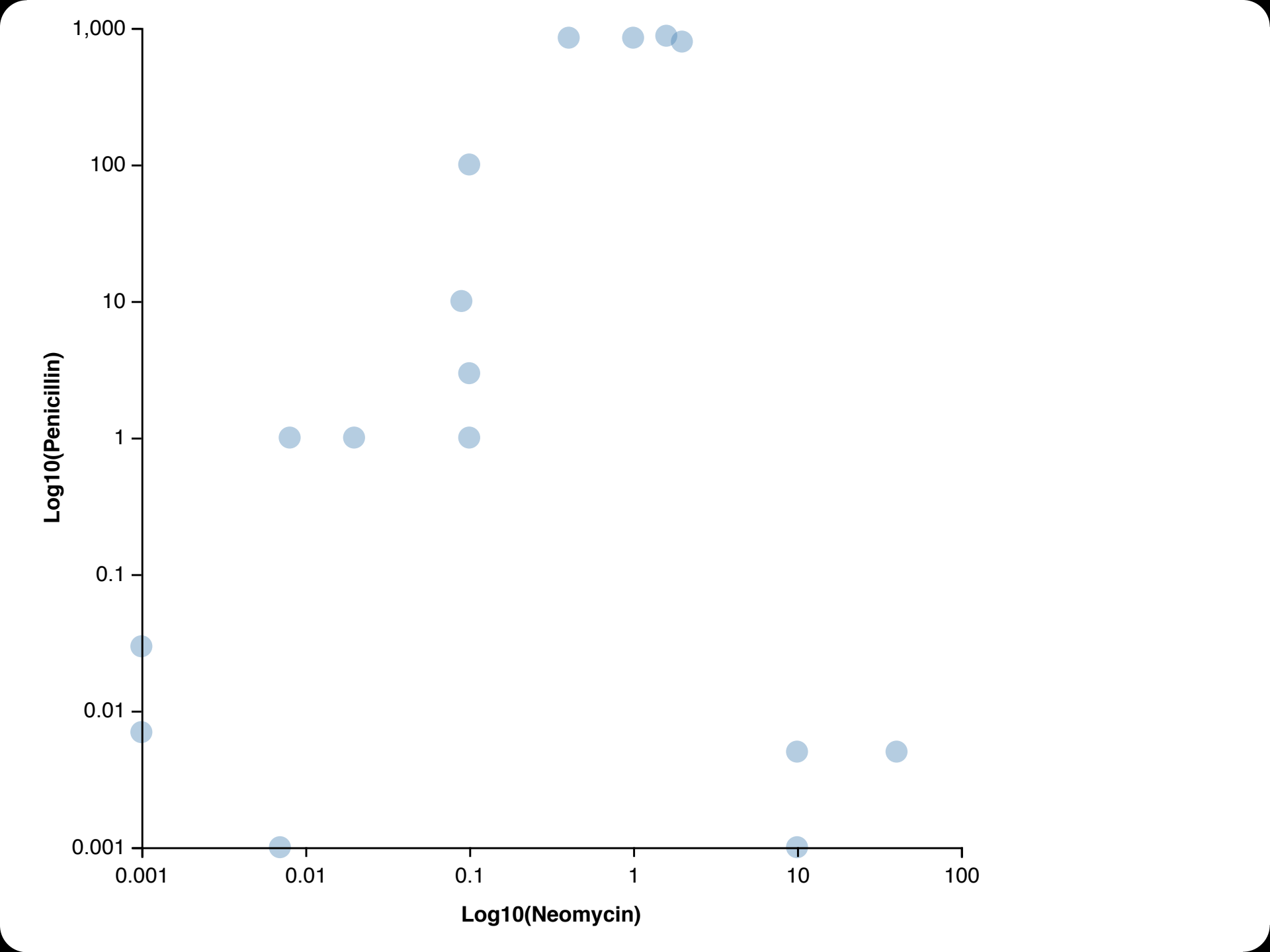

Genus
Streptococcus
Staphylococcus
Salmonella
Other
Streptococcus hemolyticus
Bacillus anthracis
Streptococcus viridans
Diplococcus pneumoniae
Staphylococcus albus
Staphylococcus aureus
Actually a streptococcus!
Streptococcus fecalis
Salmonella typhosa
Brucella abortus
Proteus vulgaris
Not a streptococcus!
Salmonella schottmuelleri
Escherichia coli
Klebsiella pneumoniae
Mycobacterium tuberculosis
Aerobacter aerogenes
Pseudomonas aeruginosa
0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
1,000
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)
Wainer & Lysen
Am. Sci. 2009

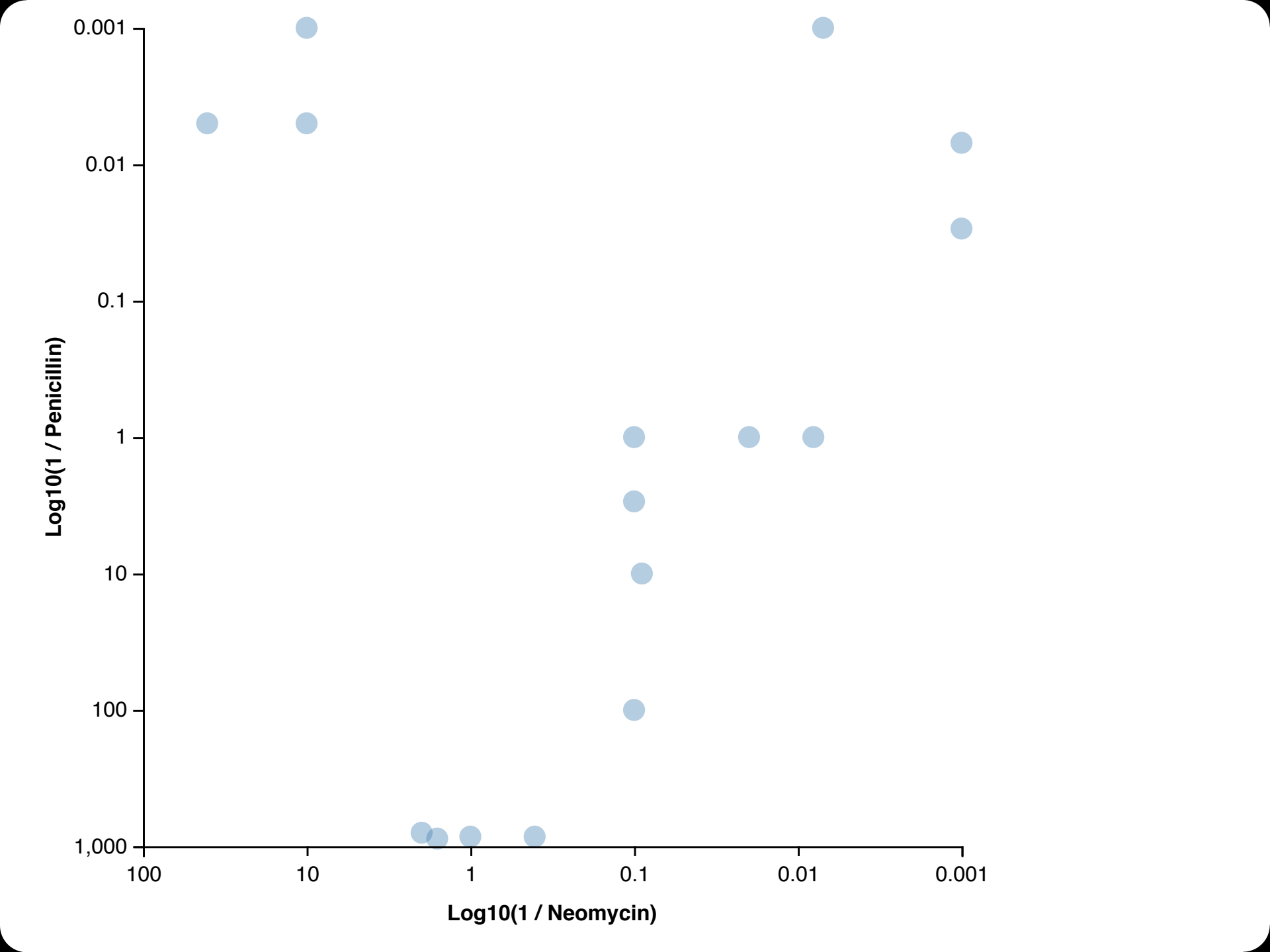
How might our tools **spur new questions** and **prompt skepticism**?

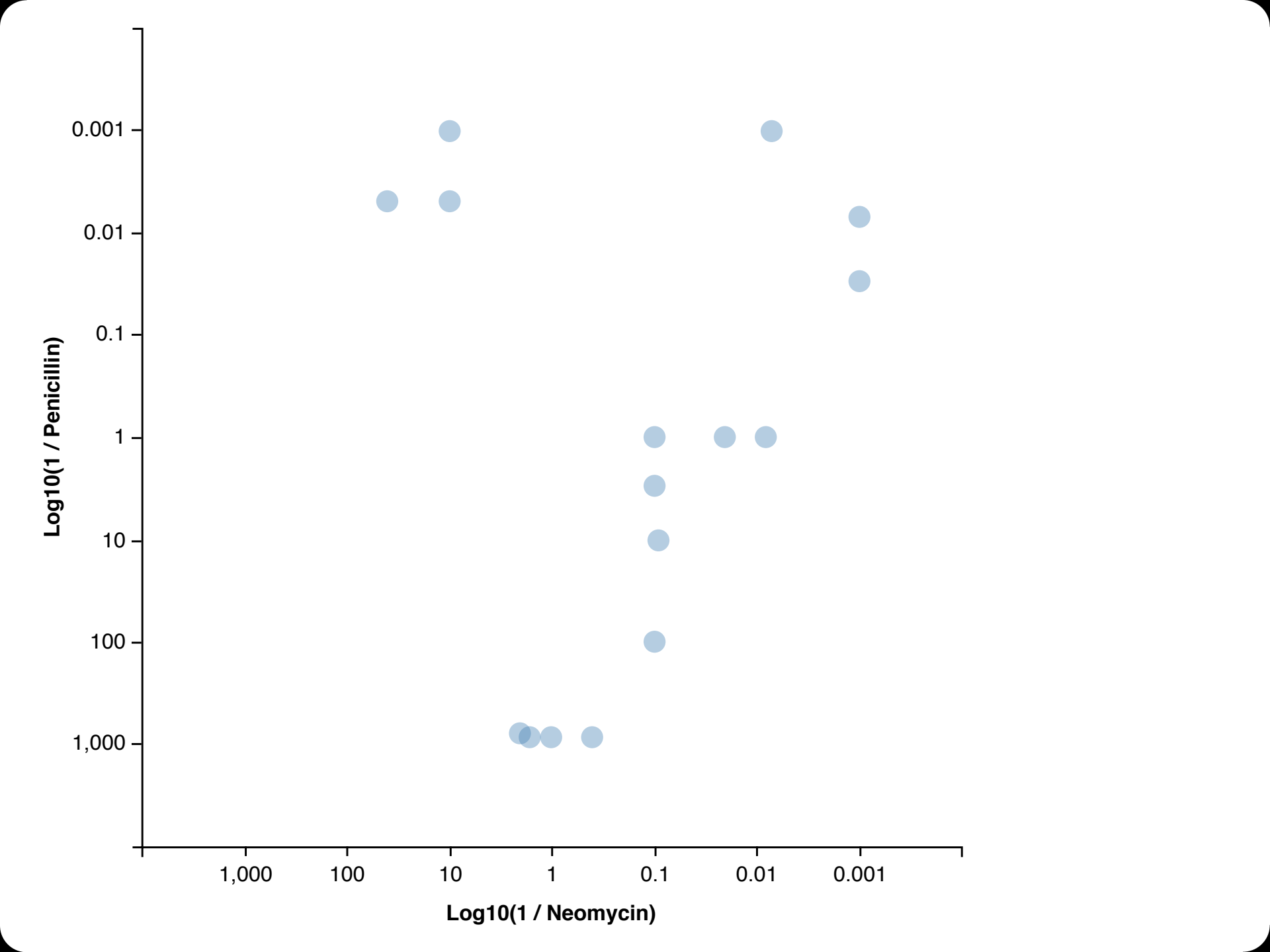
Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter aerogenes	870	1	1.6	negative
Brucella abortus	1	2	0.02	negative
Brucella anthracis	0.001	0.01	0.007	positive
Diplococcus pneumoniae	0.005	11	10	positive
Escherichia coli	100	0.4	0.1	negative
Klebsiella pneumoniae	850	1.2	1	negative
Mycobacterium tuberculosis	800	5	2	negative
Proteus vulgaris	3	0.1	0.1	negative
Pseudomonas aeruginosa	850	2	0.4	negative
Salmonella (Eberthella) typhosa	1	0.4	0.008	negative
Salmonella schottmuelleri	10	0.8	0.09	negative
Staphylococcus albus	0.007	0.1	0.001	positive
Staphylococcus aureus	0.03	0.03	0.001	positive
Streptococcus fecalis	1	1	0.1	positive
Streptococcus hemolyticus	0.001	14	10	positive
Streptococcus viridans	0.005	10	40	positive

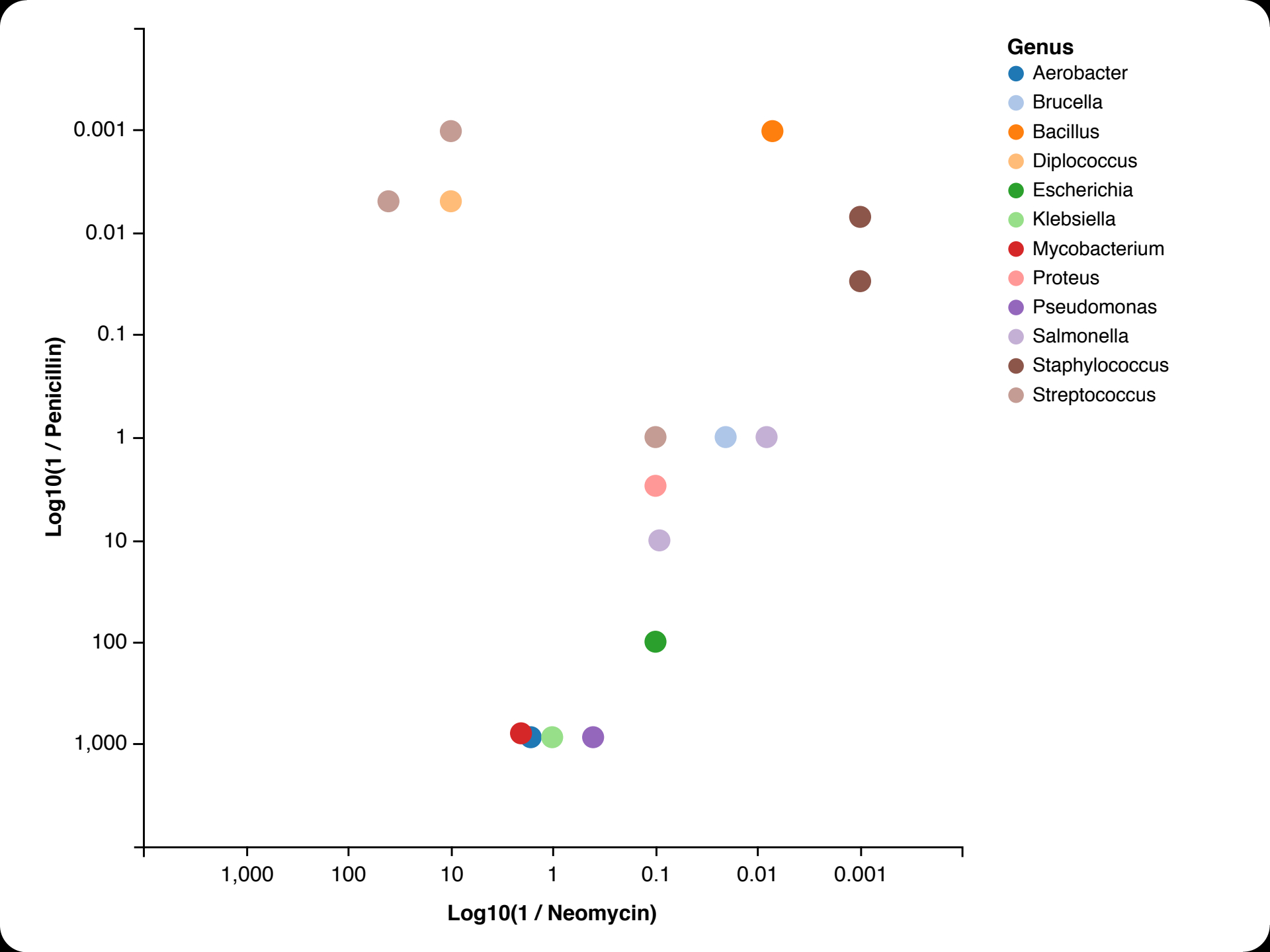
	Antibiotic			
Bacteria	**Penicillin**	Streptomycin	**Neomycin**	Gram Staining
Aerobacter aerogenes	870	1	1.6	negative
Brucella abortus	1	2	0.02	negative
Brucella anthracis	0.001	0.01	0.007	positive
Diplococcus pneumoniae	0.005	11	10	positive
Escherichia coli	100	0.4	0.1	negative
Klebsiella pneumoniae	850	1.2	1	negative
Mycobacterium tuberculosis	800	5	2	negative
Proteus vulgaris	3	0.1	0.1	negative
Pseudomonas aeruginosa	850	2	0.4	negative
Salmonella (Eberthella) typhosa	1	0.4	0.008	negative
Salmonella schottmuelleri	10	0.8	0.09	negative
Staphylococcus albus	0.007	0.1	0.001	positive
Staphylococcus aureus	0.03	0.03	0.001	positive
Streptococcus fecalis	1	1	0.1	positive
Streptococcus hemolyticus	0.001	14	10	positive
Streptococcus viridans	0.005	10	40	positive

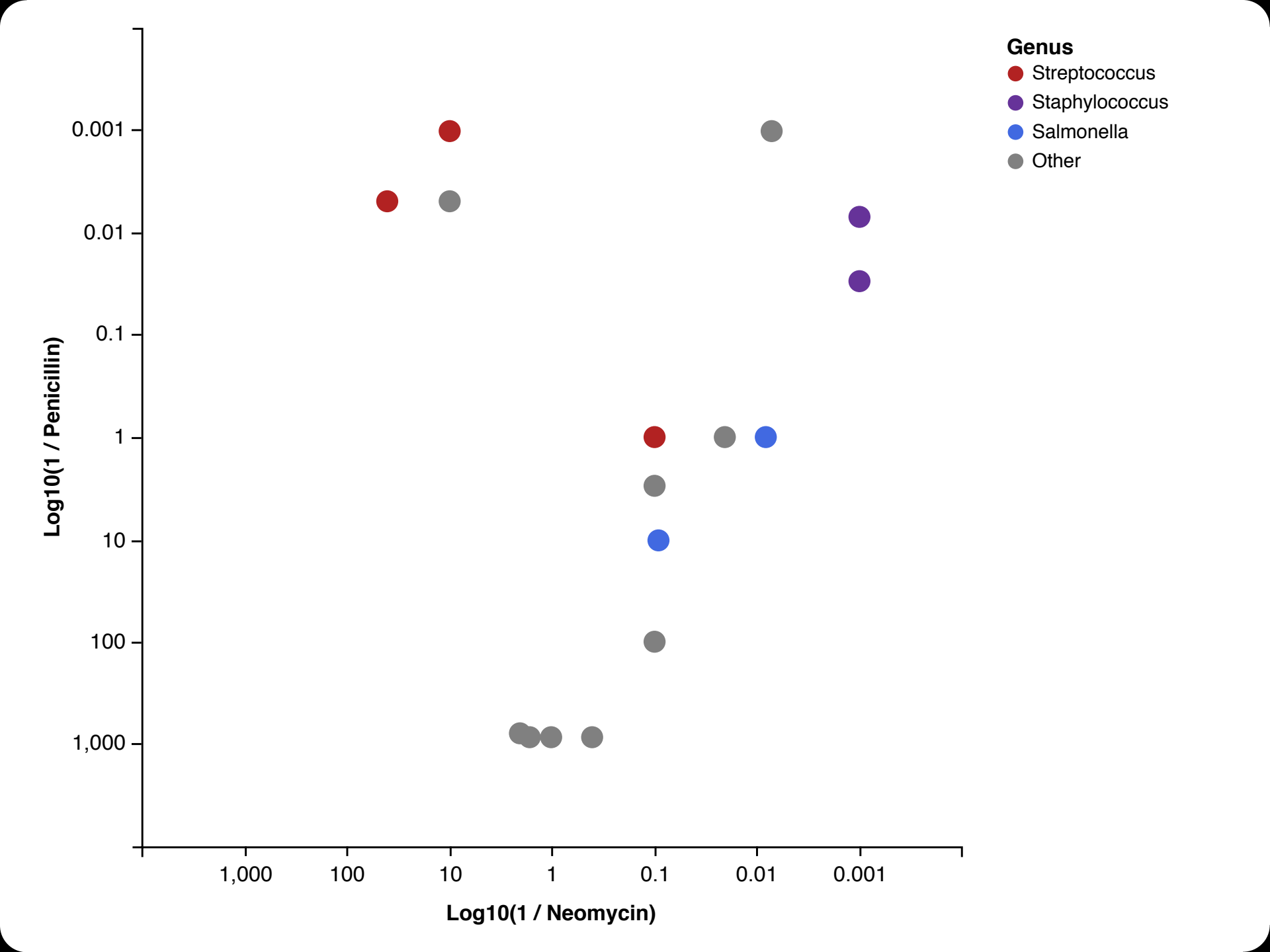

800
600
400
200
0
Penicillin
0
10
20
30
40
Neomycin

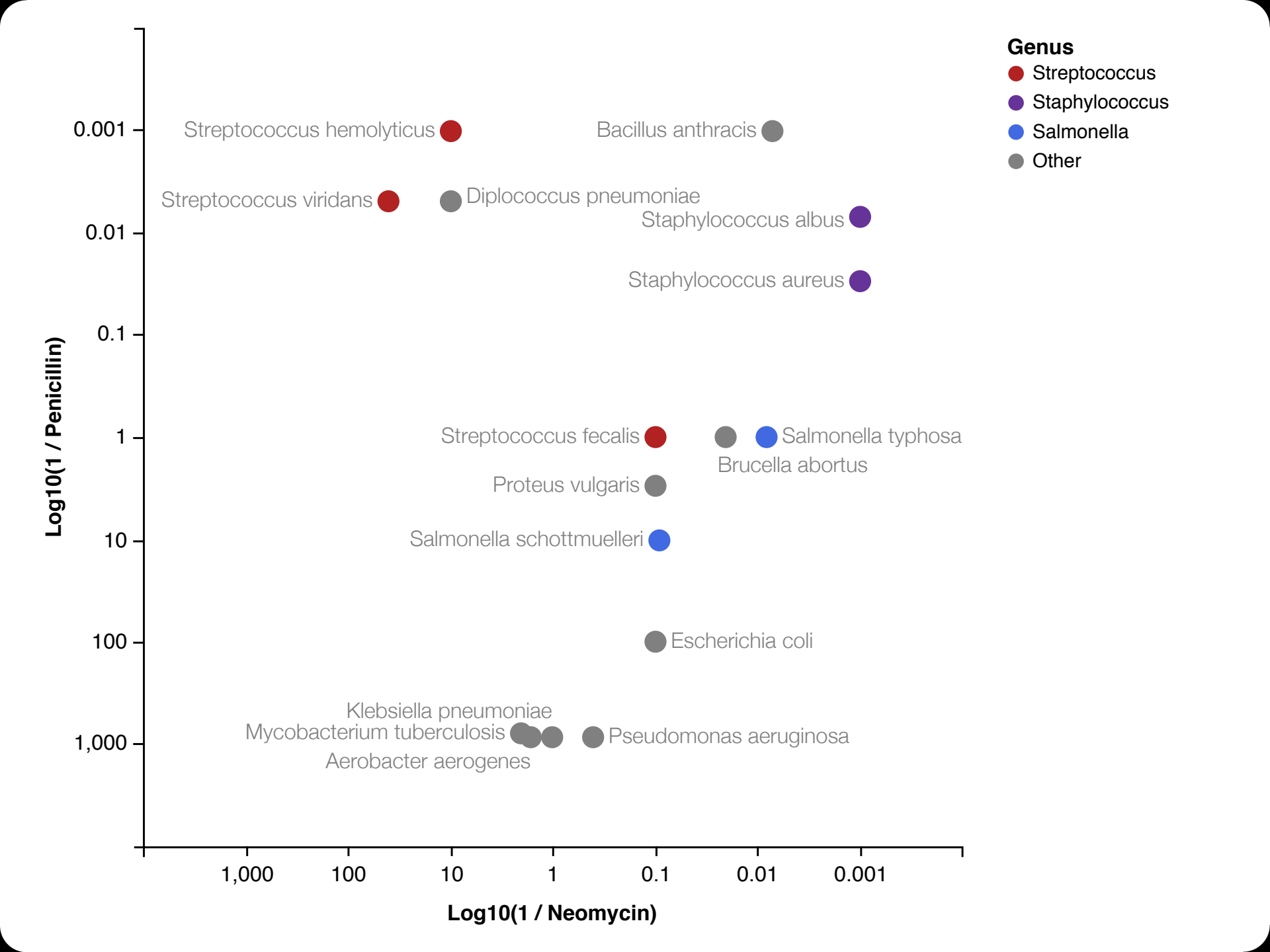

1,000
100
10
1
0.1
0.01
0.001
Log10(Penicillin)
0.001
0.01
0.1
1
10
100
Log10(Neomycin)


0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)


0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
1,000
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)


Genus
Aerobacter
Brucella
Bacillus
Diplococcus
Escherichia
Klebsiella
Mycobacterium
Proteus
Pseudomonas
Salmonella
Staphylococcus
Streptococcus
0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
1,000
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)


Genus
Streptococcus
Staphylococcus
Salmonella
Other
0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
1,000
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)


Genus
Streptococcus
Staphylococcus
Salmonella
Other
Streptococcus hemolyticus
Bacillus anthracis
Streptococcus viridans
Diplococcus pneumoniae
Staphylococcus albus
Staphylococcus aureus
Streptococcus fecalis
Salmonella typhosa
Brucella abortus
Proteus vulgaris
Salmonella schottmuelleri
Escherichia coli
Klebsiella pneumoniae
Mycobacterium tuberculosis
Aerobacter aerogenes
Pseudomonas aeruginosa
0.001
0.01
0.1
1
10
100
1,000
Log10(1 / Penicillin)
1,000
100
10
1
0.1
0.01
0.001
Log10(1 / Neomycin)

A Combinatorial Design Space

A Combinatorial Design Space

1. Variable Selection

A Combinatorial Design Space

1. Variable Selection
2. Data Transformation

A Combinatorial Design Space

1. Variable Selection
2. Data Transformation
3. Visual Encoding Design

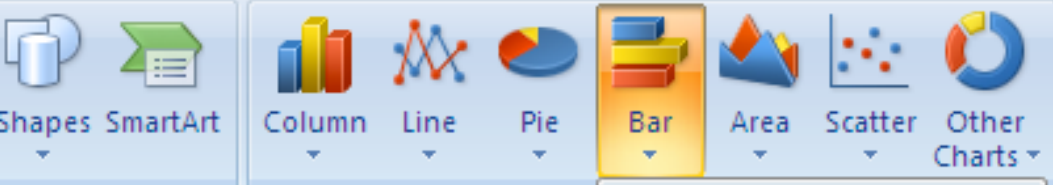
A Combinatorial Design Space

1. Variable Selection
2. Data Transformation
3. Visual Encoding Design

—> *Thousands of possible charts!*

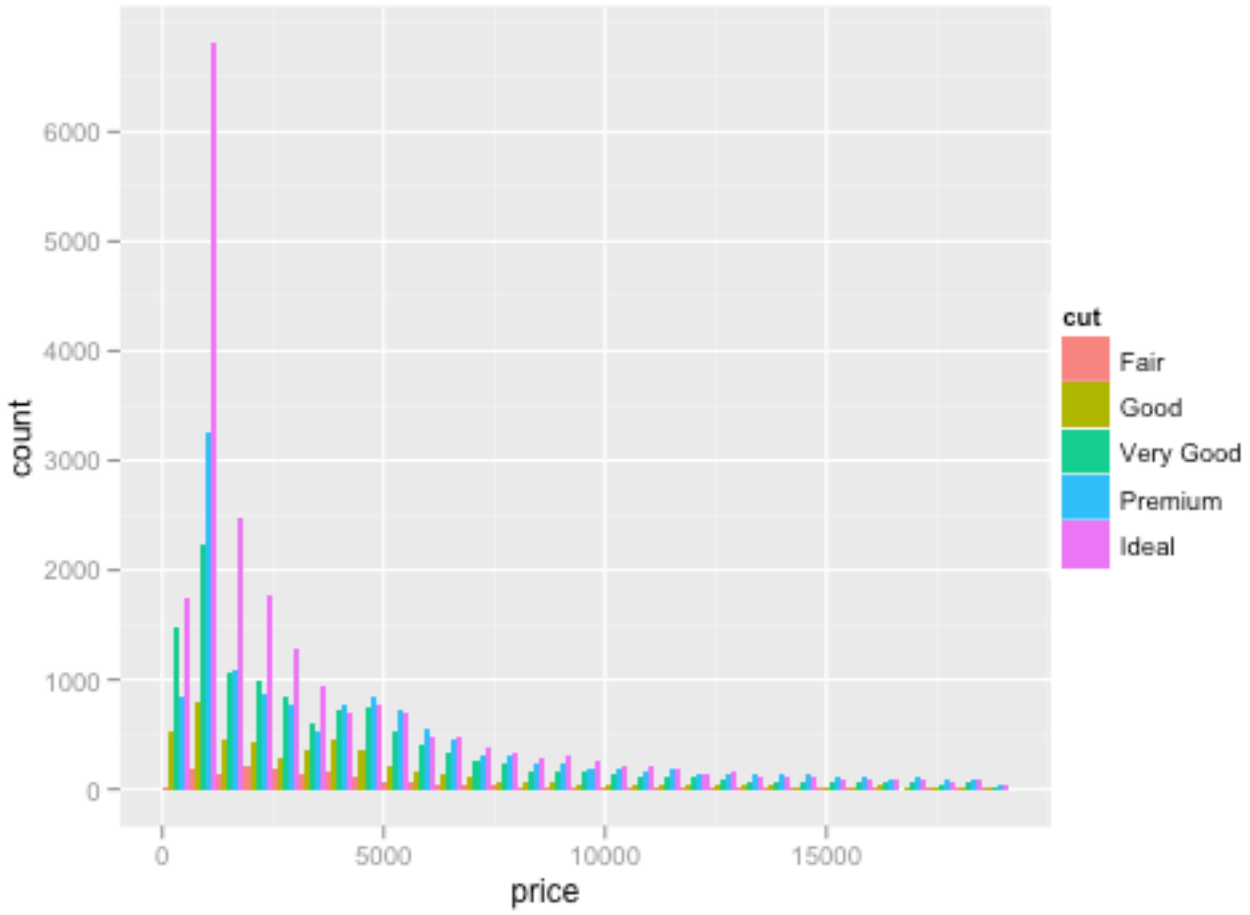
How might we augment manual chart construction with **interactive browsing of recommended visualizations**?

Layout Formulas Data Review View Add-Ins



fx L1

B	C	D		H	I	J
	task	row	col	d2	L1	L2
-color	SA	1		0	0	0
-color	SA	1		0	0.73568	0.73568
-color	SA	1		0	0.73119	0.73119
-color	SA	1		0	0.24612	0.24612
-color	SA	1		0.8119	0.8119	0.8119
-color	SA	1		0.8119	1.54758	1.095630719
-color	SA	1		0.8119	1.54309	1.092620898
-color	SA	1		0.8119	1.05802	0.848384738
-color	SA	1		0.8417	0.8417	0.8417
-color	SA	1		0.8417	1.57738	1.117892639
-color	SA	1		0.8417	1.57289	1.114942916
-color	SA	1		0.8417	1.08782	0.876945805
-color	SA	1		0.2825	0.2825	0.2825
-color	SA	1		0.2825	1.01818	0.788055399
-color	SA	1		0.2825	1.01369	0.783865464
-color	SA	1	16	0.2825	0.52862	0.374674932




```
ggplot(diamonds, aes(x=price, fill=cut))
+ geom_bar(position="dodge")
```

```
var svg = div.append("svg:svg")
    .attr("width", w)
    .attr("height", w)
  .append("svg:g")
    .attr("transform", "translate(" + rx + "," + ry + ")");

svg.append("svg:path")
    .attr("class", "arc")
    .attr("d", d3.svg.arc().outerRadius(ry - 120).innerRadius(0).startAngle(0).endAngle(2 * Math.PI))
    .on("mousedown", mousedown);

d3.json("data/flare-imports.json", function(classes) {
  var nodes = cluster.nodes(packages.root(classes)),
      links = packages.imports(nodes),
      splines = bundle(links);

  var path = svg.selectAll("path.link")
      .data(links)
    .enter().append("svg:path")
      .attr("class", function(d) { return "link source-" + d.source.key + " target-" + d.target.key; })
      .attr("d", function(d, i) { return line(splines[i]); });

  svg.selectAll("g.node")
      .data(nodes.filter(function(n) { return !n.children; }))
    .enter().append("svg:g")
      .attr("class", "node")
      .attr("id", function(d) { return "node-" + d.key; })
      .attr("transform", function(d) { return "rotate(" + (d.x - 90) + ")translate(" + d.y + ")"; })
    .append("svg:text")
      .attr("dx", function(d) { return d.x < 180 ? 8 : -8; })
      .attr("dy", ".31em")
      .attr("text-anchor", function(d) { return d.x < 180 ? "start" : "end"; })
      .attr("transform", function(d) { return d.x < 180 ? null : "rotate(180)"; })
      .text(function(d) { return d.key; })
      .on("mouseover", mouseover)
      .on("mouseout", mouseout);

  d3.select("input[type=range]").on("change", function() {
    line.tension(this.value / 100);
    path.attr("d", function(d, i) { return line(splines[i]); });
  });
});
```

1. Trifacta Visual Profiler

2. Data Voyager (UW + Tableau)

1. **Trifacta Visual Profiler**

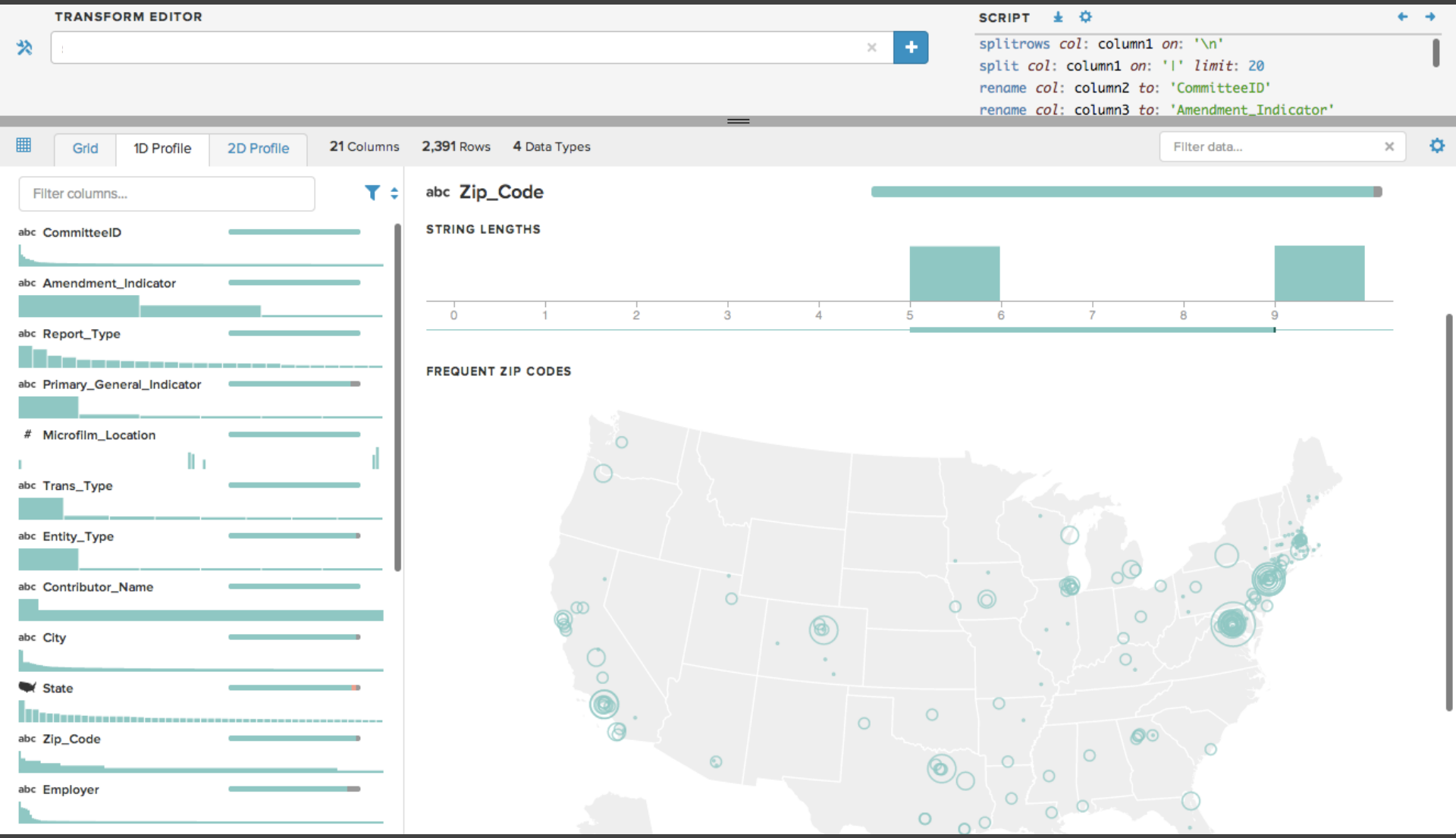
2. Data Voyager (UW + Tableau)

Trifacta Visual Profiler

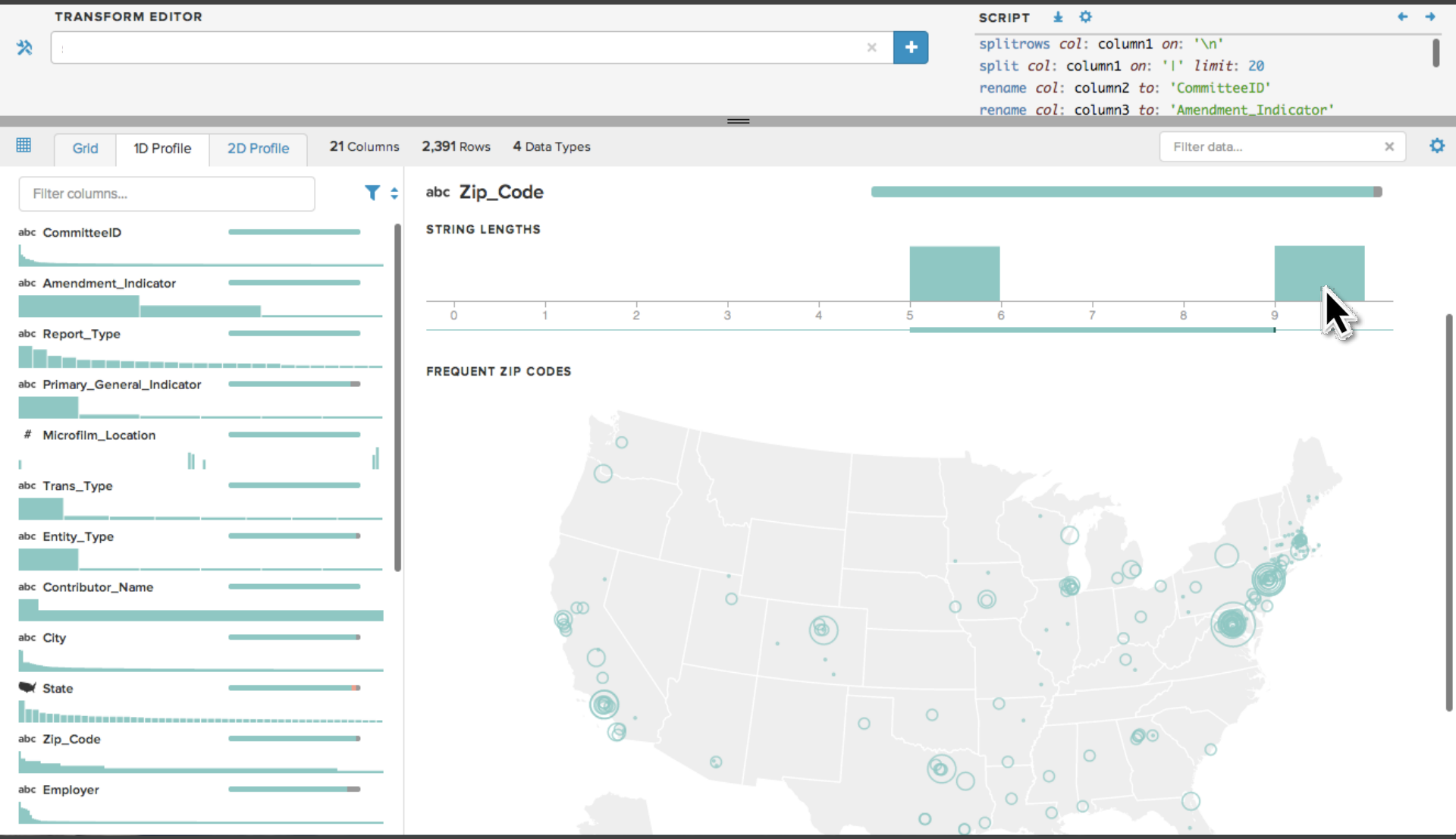
Trifacta Visual Profiler

Details

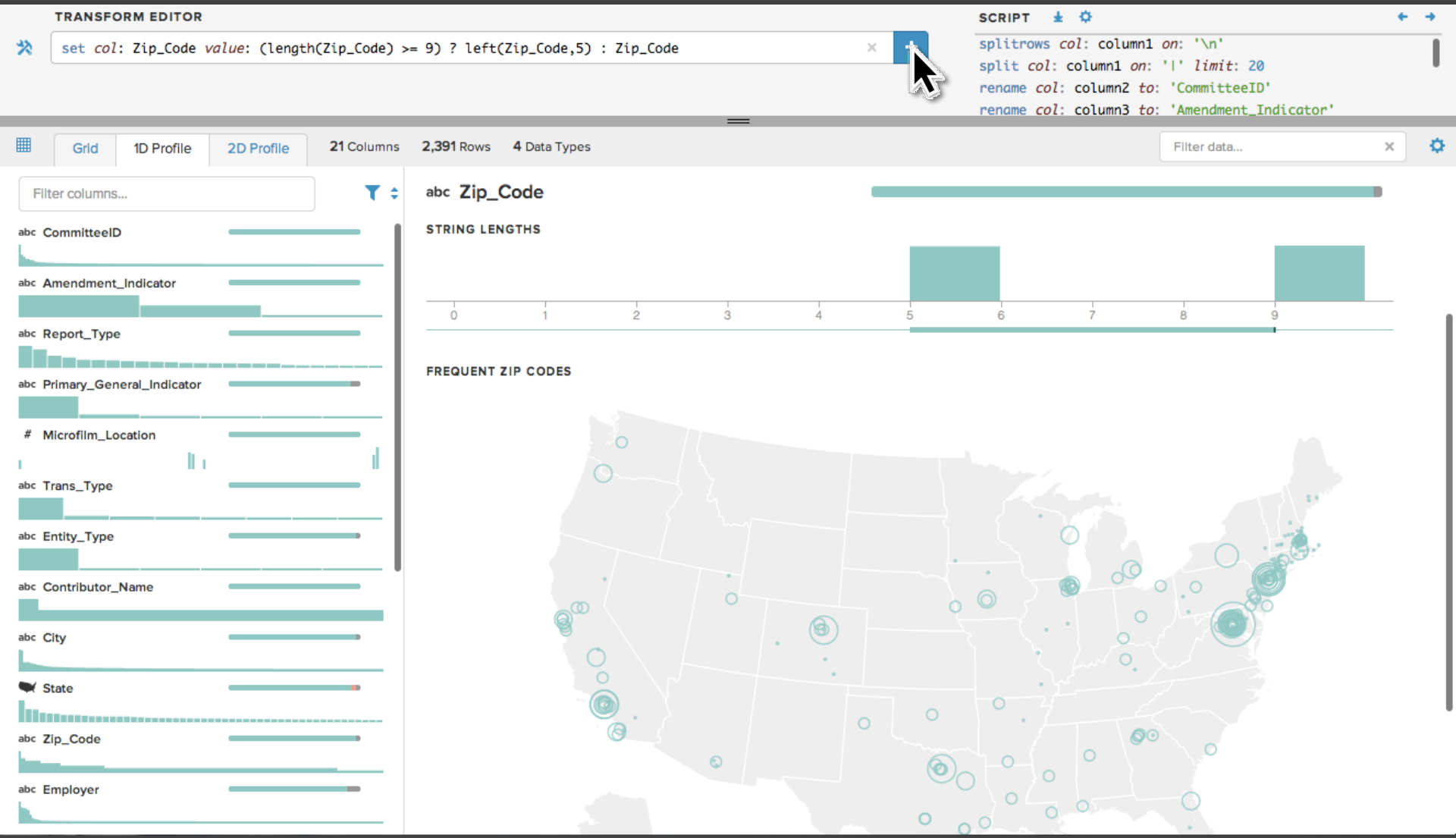
Trifacta Visual Profiler



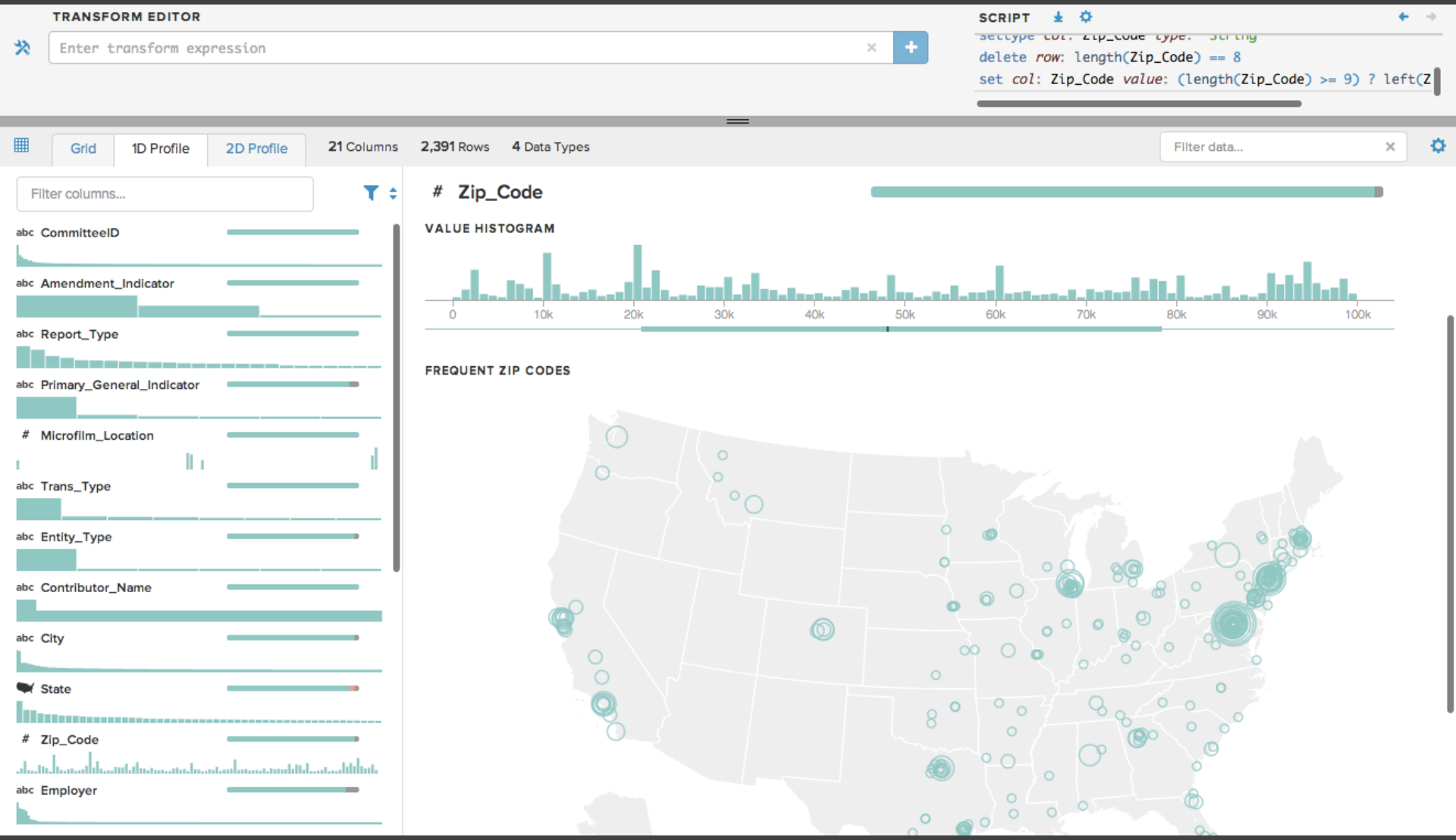
Trifacta Visual Profiler



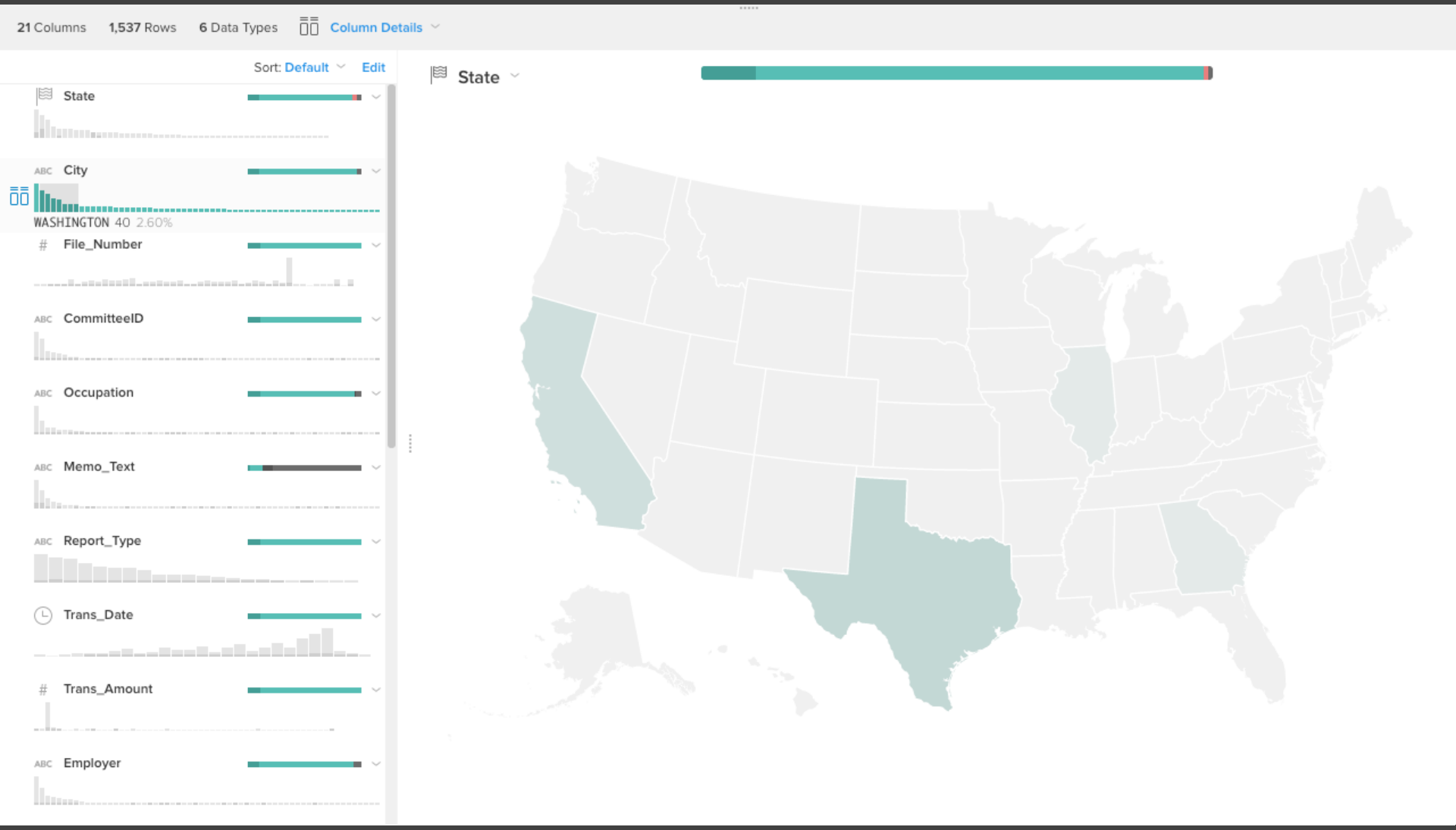
Trifacta Visual Profiler



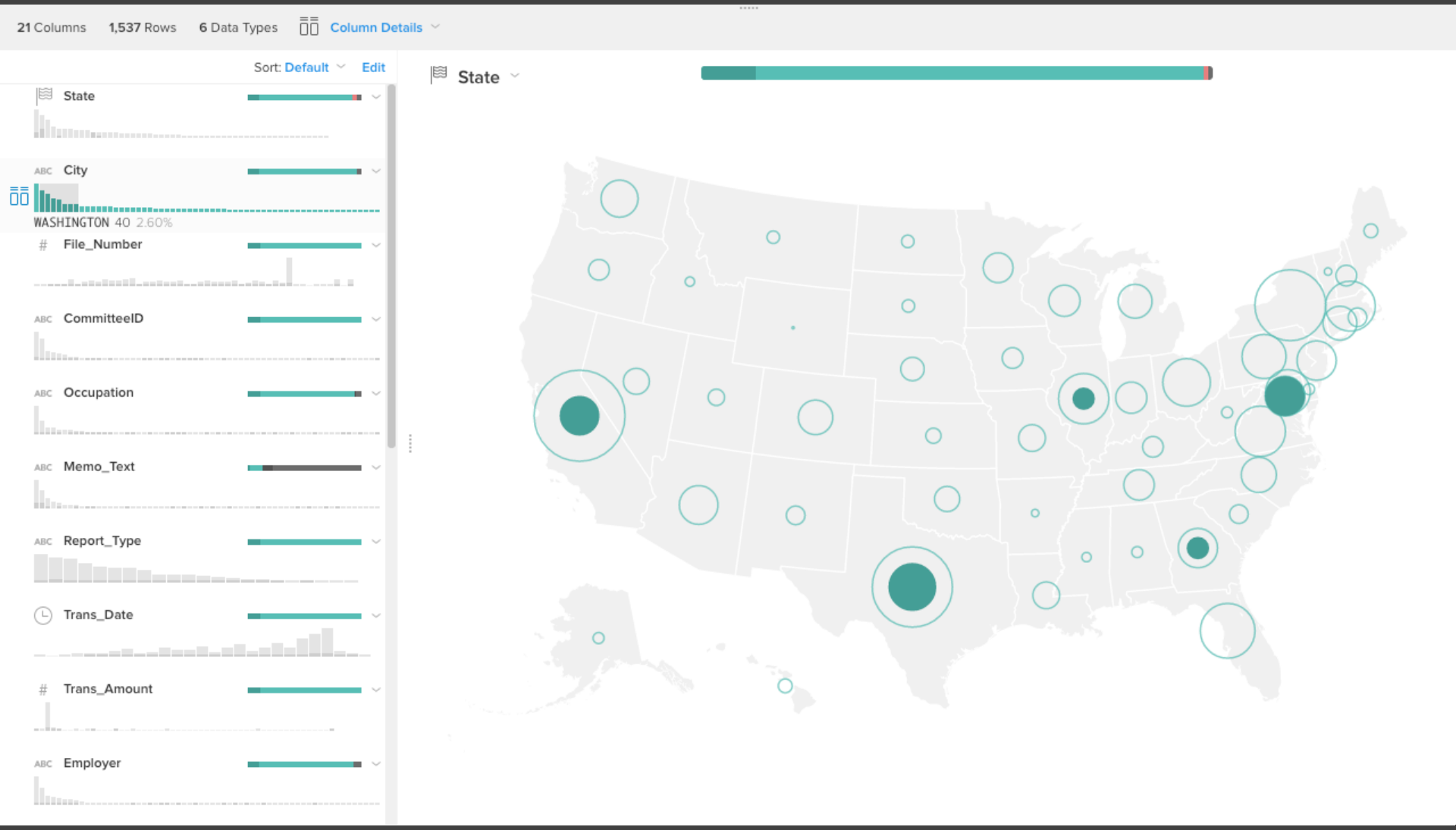
Trifacta Visual Profiler



Trifacta Visual Profiler



Trifacta Visual Profiler



Trifacta Visual Profiler

1. Trifacta Visual Profiler

2. **Data Voyager** (UW + Tableau)

Data Voyager

https://uwdata.github.io/voyager

Data Voyager

Bookmarks (0) Undo Redo

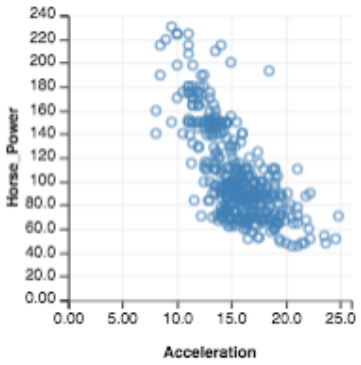
DATA Cars

- Cylinders
- name
- Origin
- AUTO (Year)
- AUTO (Acceleration)
- AUTO (Displacement)
- AUTO (Horse Power)
- AUTO (Miles per Gallon)
- AUTO (Weight in lbs)
- COUNT

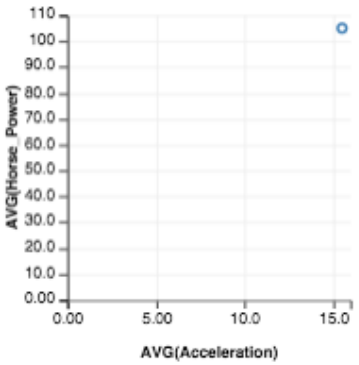
Reset

Showing Data Variations for # AUTO (Acceleration) # AUTO (Horse Power)

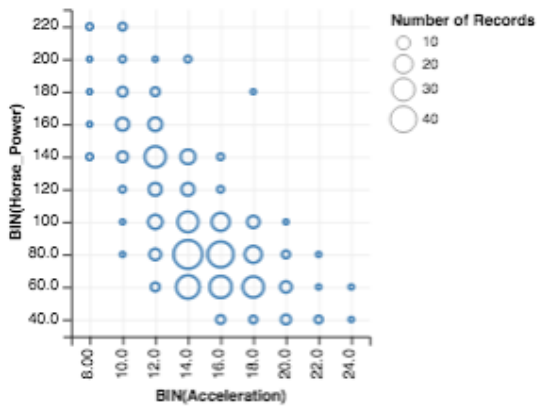
Acceleration # Horse Power



AVG (Acceleration) # AVG (Horse Power)

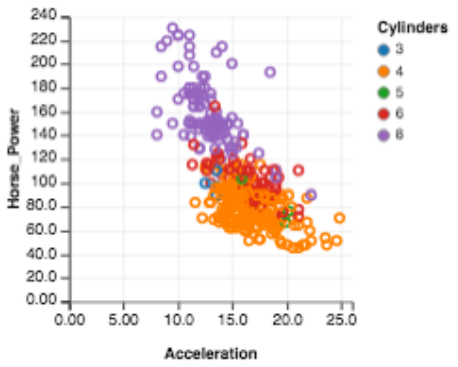


BIN (Acceleration) # BIN (Horse Power) # COUNT

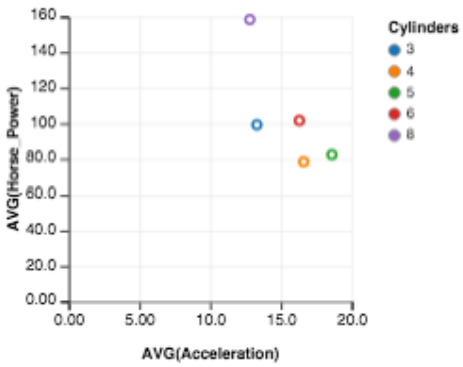


Showing Data Variations for # AUTO (Acceleration) # AUTO (Horse Power) **with one additional field.**

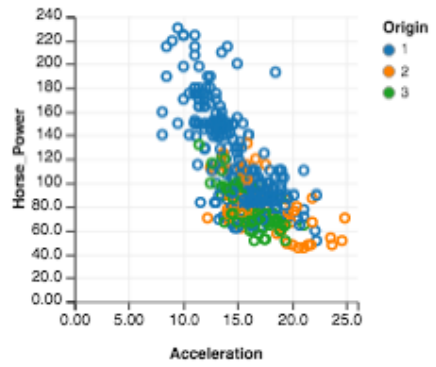
Acceleration # Horse Power Abc Cylinders



AVG (Acceleration) # AVG (Horse Power) Abc Cylinders



Acceleration # Horse Power Abc Origin



AVG (Acceleration) # AVG (Horse Power) Abc Origin

Acceleration # Horse Power YEAR (Year)

User


Data Set
Voyager
Visualization Browser
User

User

Compass

Recommendation Engine

Data Schema
& Statistics

Voyager

Visualization Browser

1. Select **data variables**
2. Apply **transformations**
3. Pick visual **encodings**

User

Compass

Recommendation Engine

Data Schema
& Statistics

Voyager

Visualization Browser

Constrain & rank choices by **data type**, **statistics** & **perceptual principles.**

User

Compass

Recommendation Engine

Data Schema & Statistics

Voyager

Visualization Browser

User

Compass

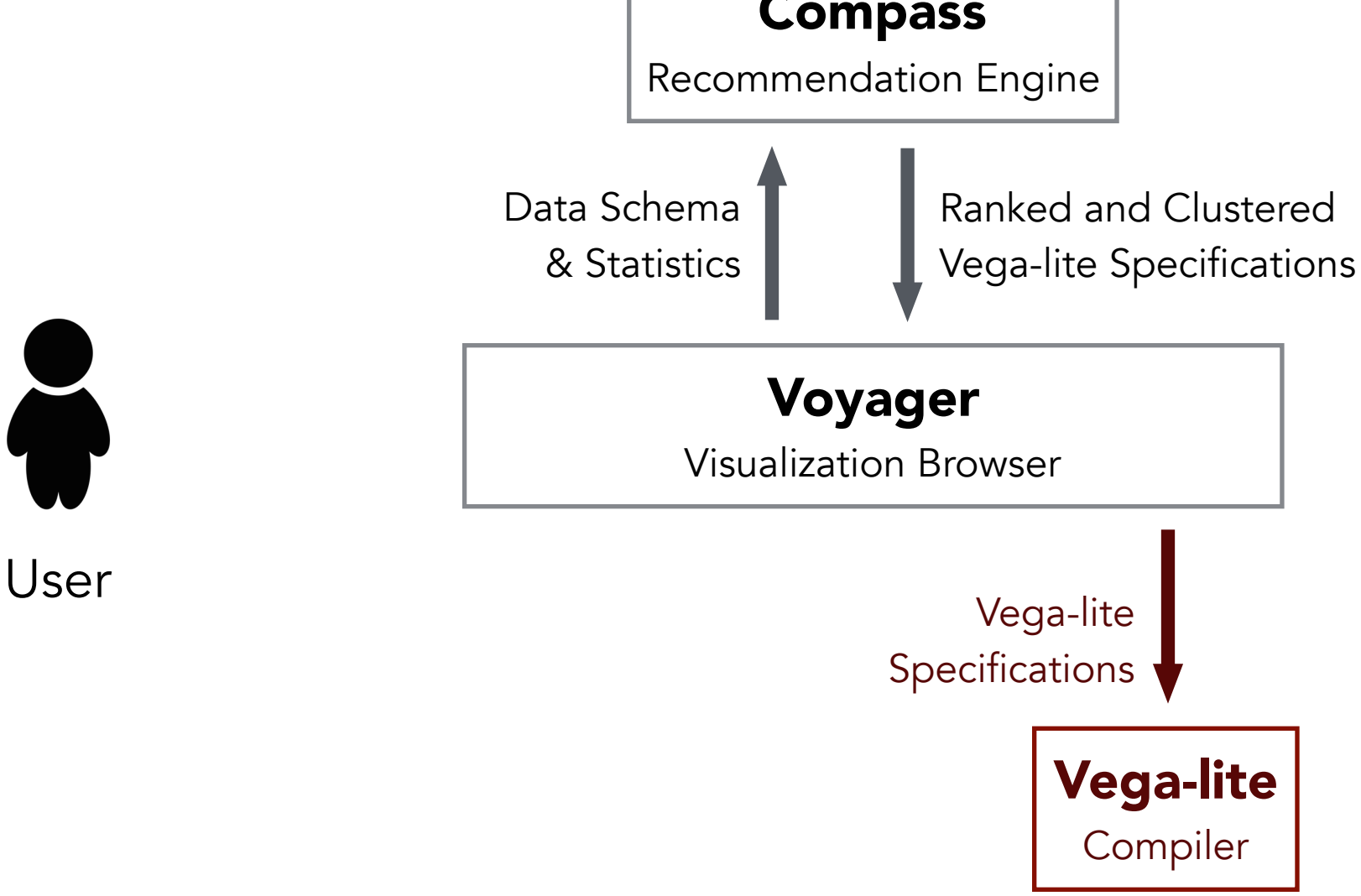
Recommendation Engine

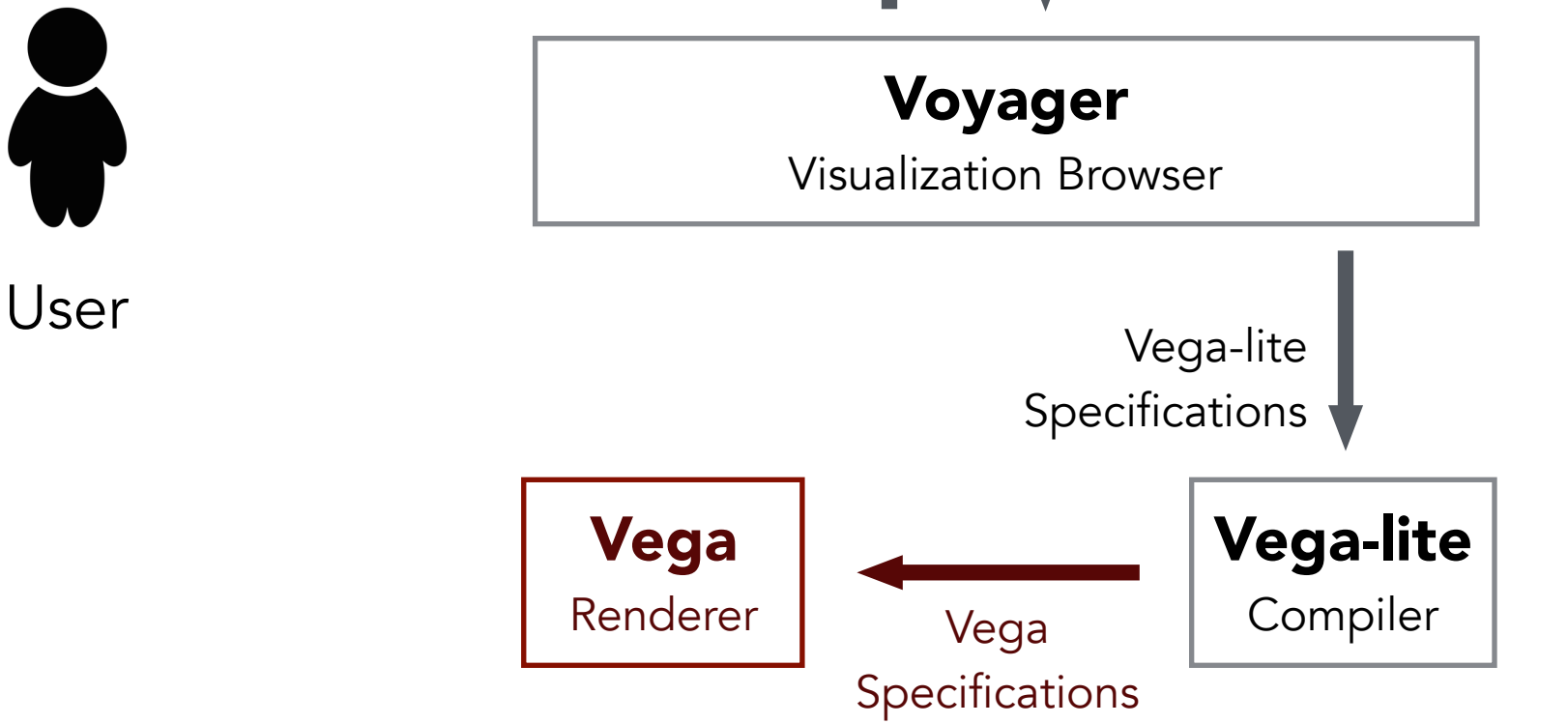
Data Schema
& Statistics

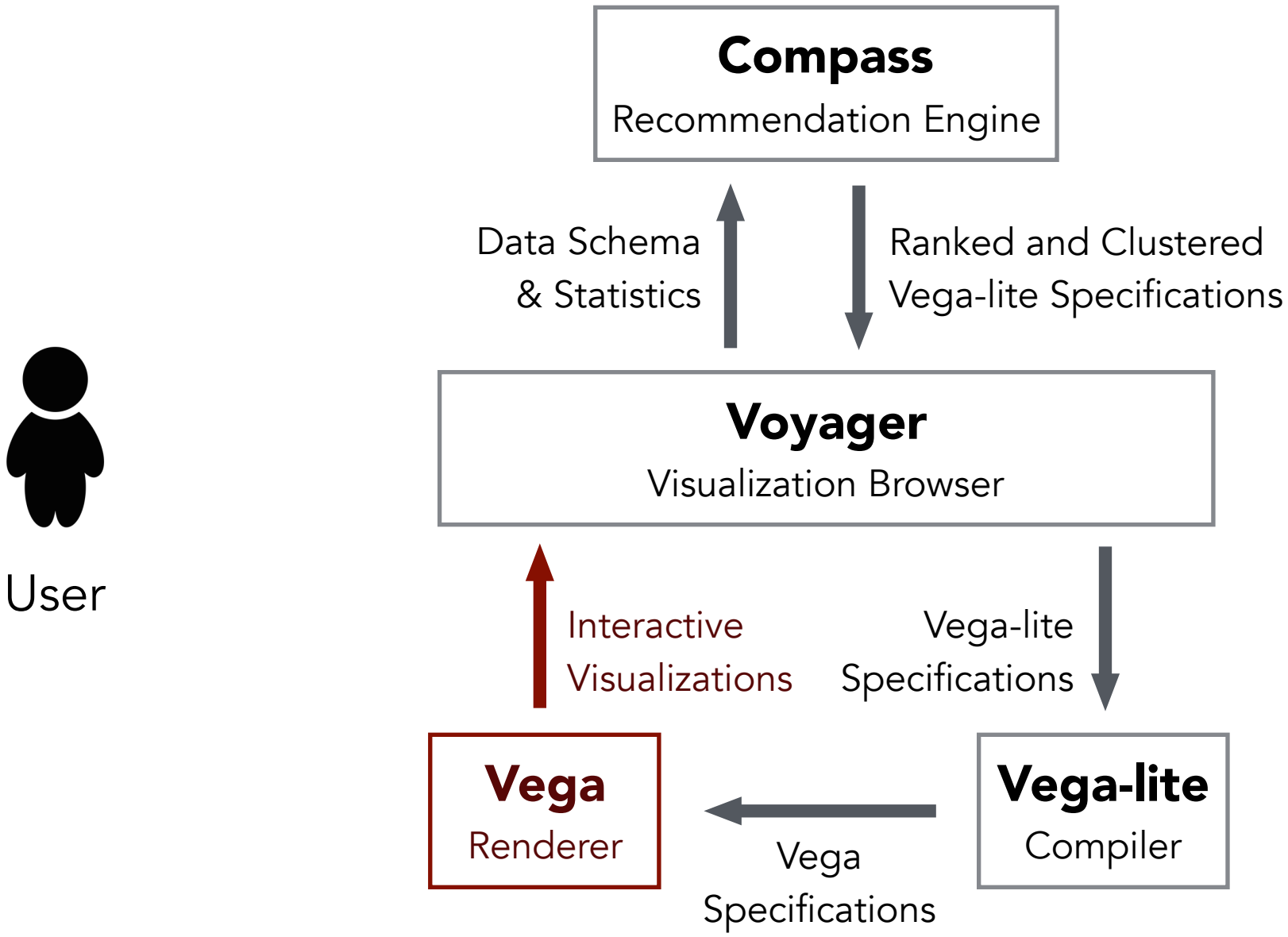
Ranked and Clustered
Vega-lite Specifications

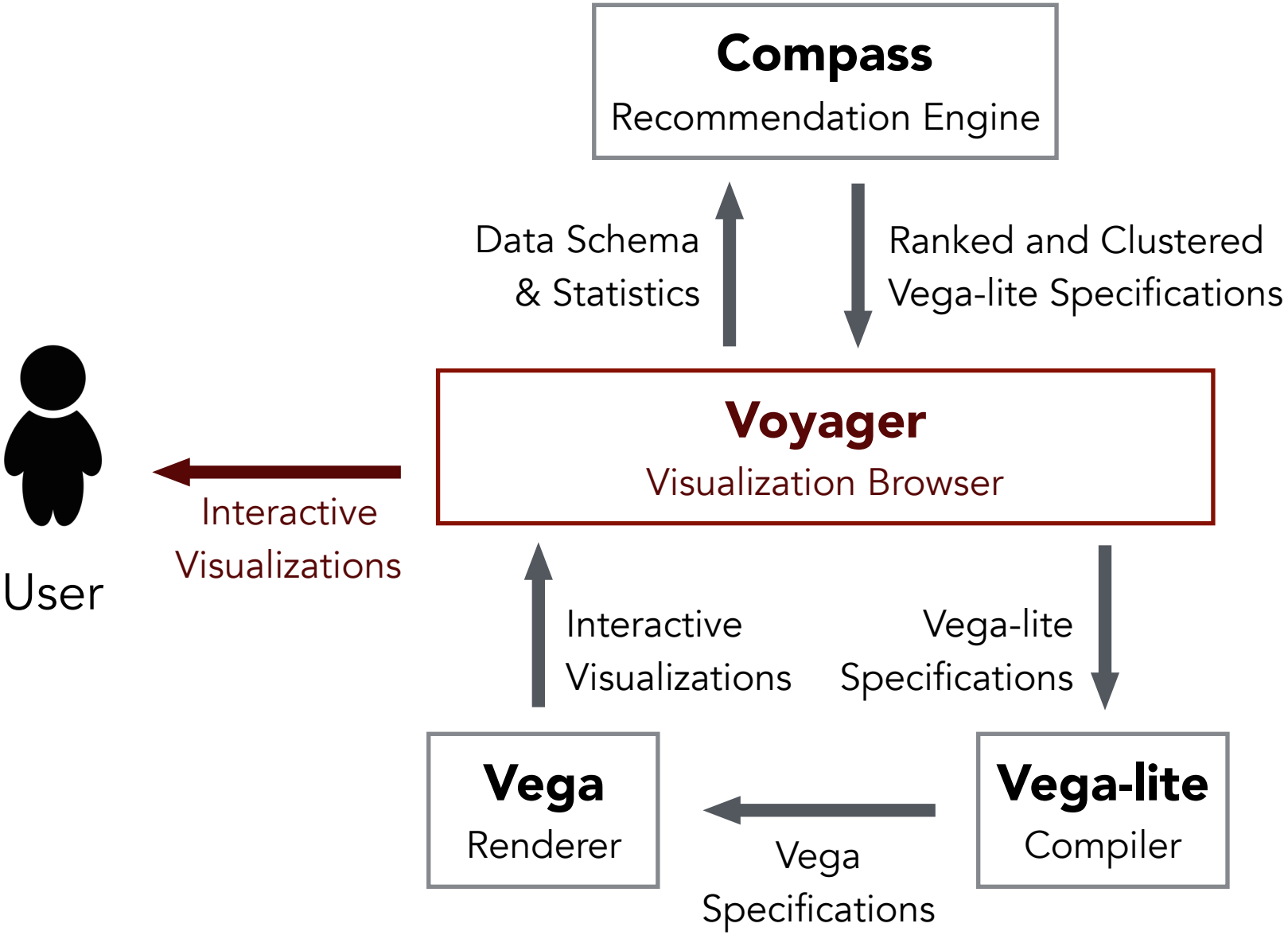
Voyager

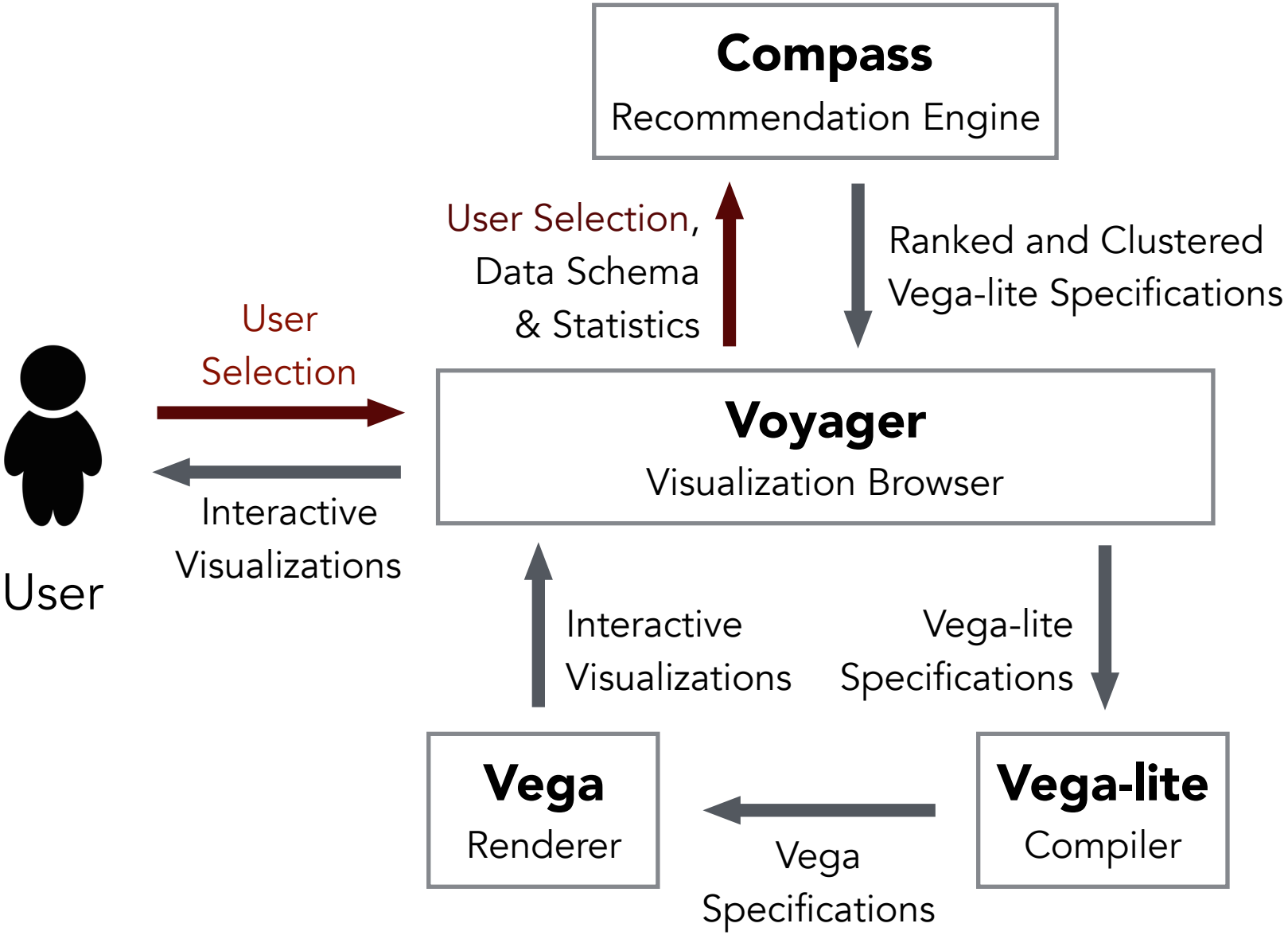
Visualization Browser


Compass
Recommendation Engine
Data Schema
& Statistics
Ranked and Clustered
Vega-lite Specifications
Voyager
Visualization Browser
User
Vega-lite
Specifications
Vega-lite
Compiler


Compass
Recommendation Engine
Data Schema
& Statistics
Ranked and Clustered
Vega-lite Specifications
Voyager
Visualization Browser
User
Vega-lite
Specifications
Vega
Renderer
Vega
Specifications
Vega-lite
Compiler


Compass
Recommendation Engine
Data Schema
& Statistics
Ranked and Clustered
Vega-lite Specifications
Voyager
Visualization Browser
User
Interactive
Visualizations
Vega-lite
Specifications
Vega
Renderer
Vega
Specifications
Vega-lite
Compiler

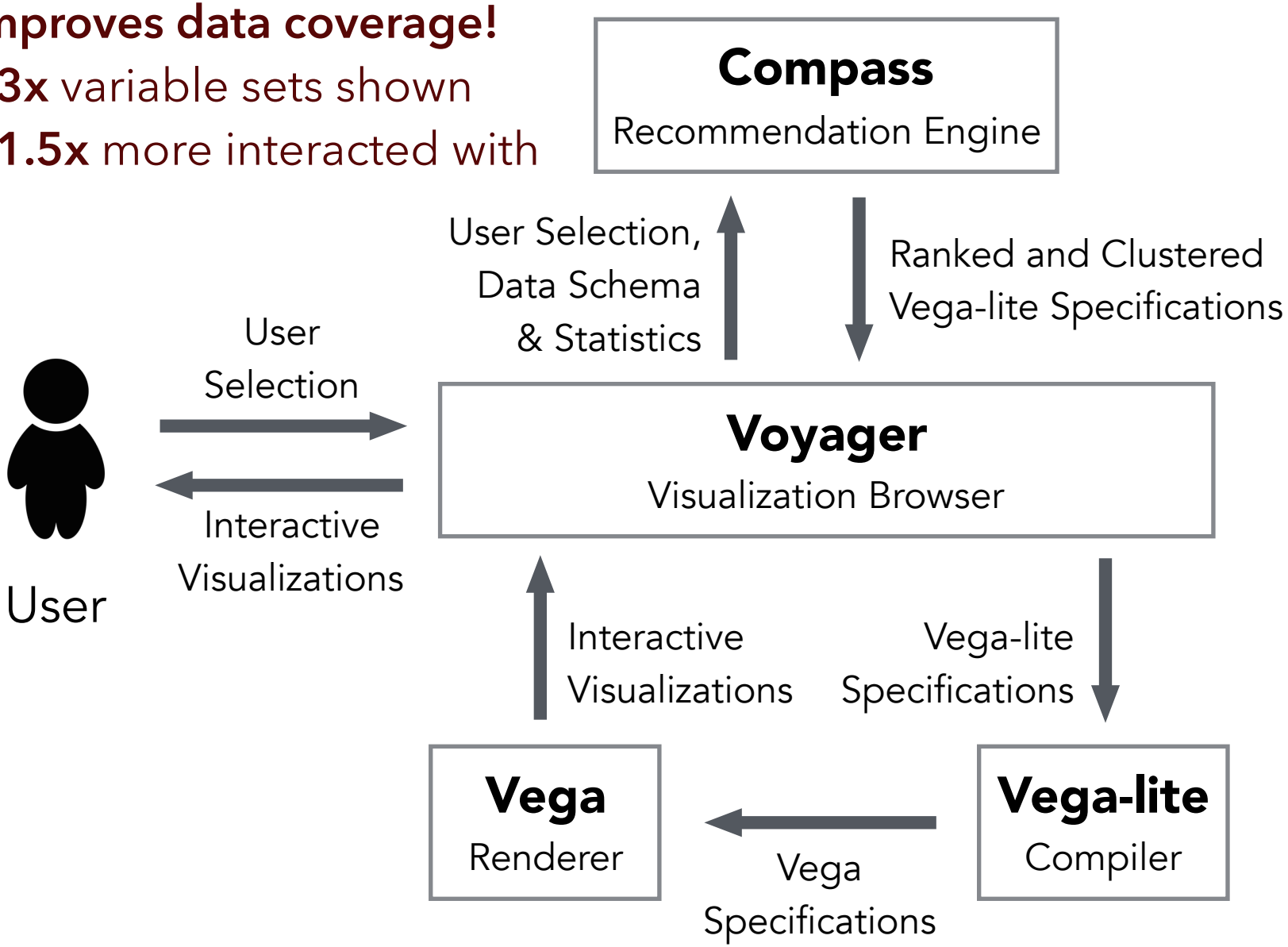

Compass
Recommendation Engine
Data Schema
& Statistics
Ranked and Clustered
Vega-lite Specifications
Voyager
Visualization Browser
User
Interactive
Visualizations
Interactive
Visualizations
Vega-lite
Specifications
Vega
Renderer
Vega-lite
Compiler
Vega
Specifications


Compass
Recommendation Engine
User Selection,
Data Schema
& Statistics
Ranked and Clustered
Vega-lite Specifications
User
Selection
Interactive
Visualizations
Voyager
Visualization Browser
User
Interactive
Visualizations
Vega-lite
Specifications
Vega
Renderer
Vega
Specifications
Vega-lite
Compiler

Improves data coverage!

+**3x** variable sets shown

+**1.5x** more interacted with



Ongoing Challenges

Refining visualization recommendation
What to optimize? How to evaluate?

Scaling interactive visualizations
Large *D* harder than large *N*…

Help avoid statistical pitfalls
Recognize mix effects, convey uncertainty

How might our tools **spur new questions** and **prompt skepticism**?

ISTC BIG DATA

DARPA

GORDON AND BETTY MOORE FOUNDATION

Vega

https://vega.github.io

Vega A VISUALIZATION GRAMMAR

Vega is a declarative format for creating, saving, and sharing visualization designs. With Vega, visualizations are described in JSON, and generate interactive views using either HTML5 Canvas or SVG.

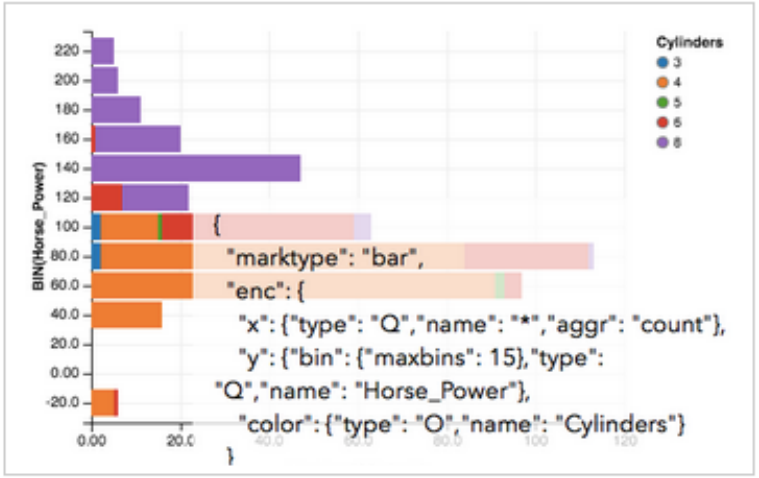
TOOLKITS

VEGA offers a full declarative visualization grammar, suitable for expressive custom interactive visualization design and programmatic generation.

Tutorial | Documentation | Discussion Forum

v1.5 (stable): download, examples, github

NEW v2.0 (dev): download, examples, github



NEW **VEGALITE** provides a higher-level grammar for visual analysis, comparable to ggplot or Tableau, that generates complete Vega specifications.

Online Editor | GitHub

vega.github.io

Visualization for DISCOVERY

Jeffrey Heer @jeffrey_heer
Univ. of Washington + Trifacta