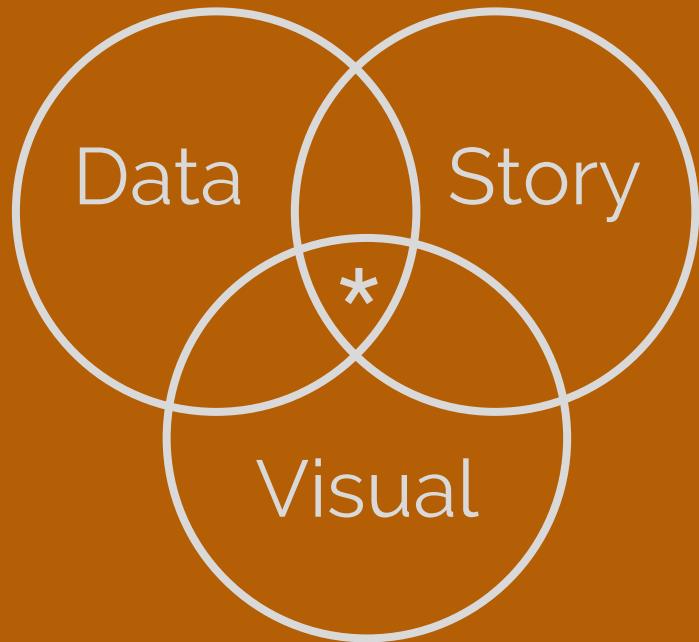


# Data Visualisation

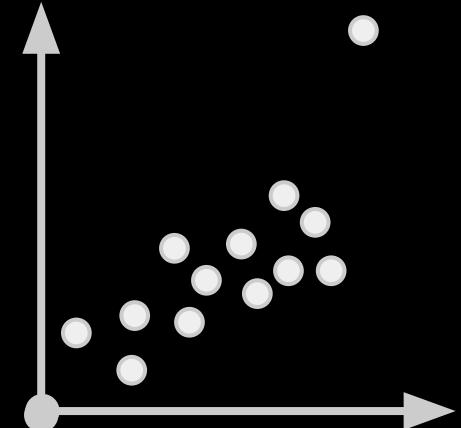
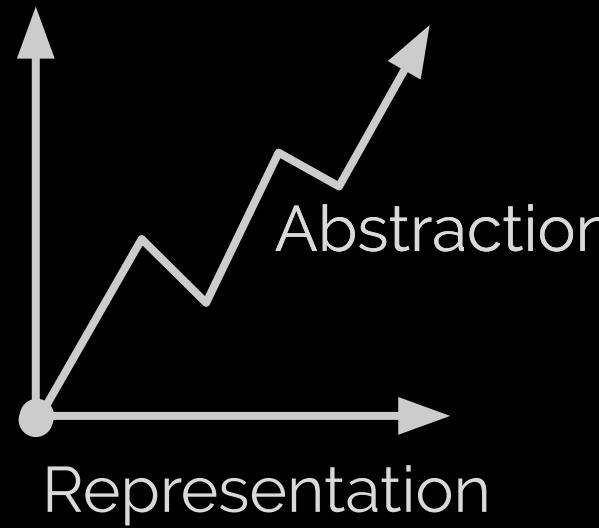
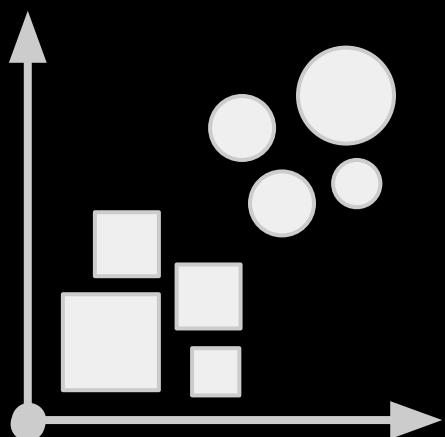
Amit Kapoor  
@amitkaps

Representation



# 2

# Representation



# Visualise Large Data



Pin codes  
in India

~24,000 Pin codes

e.g.

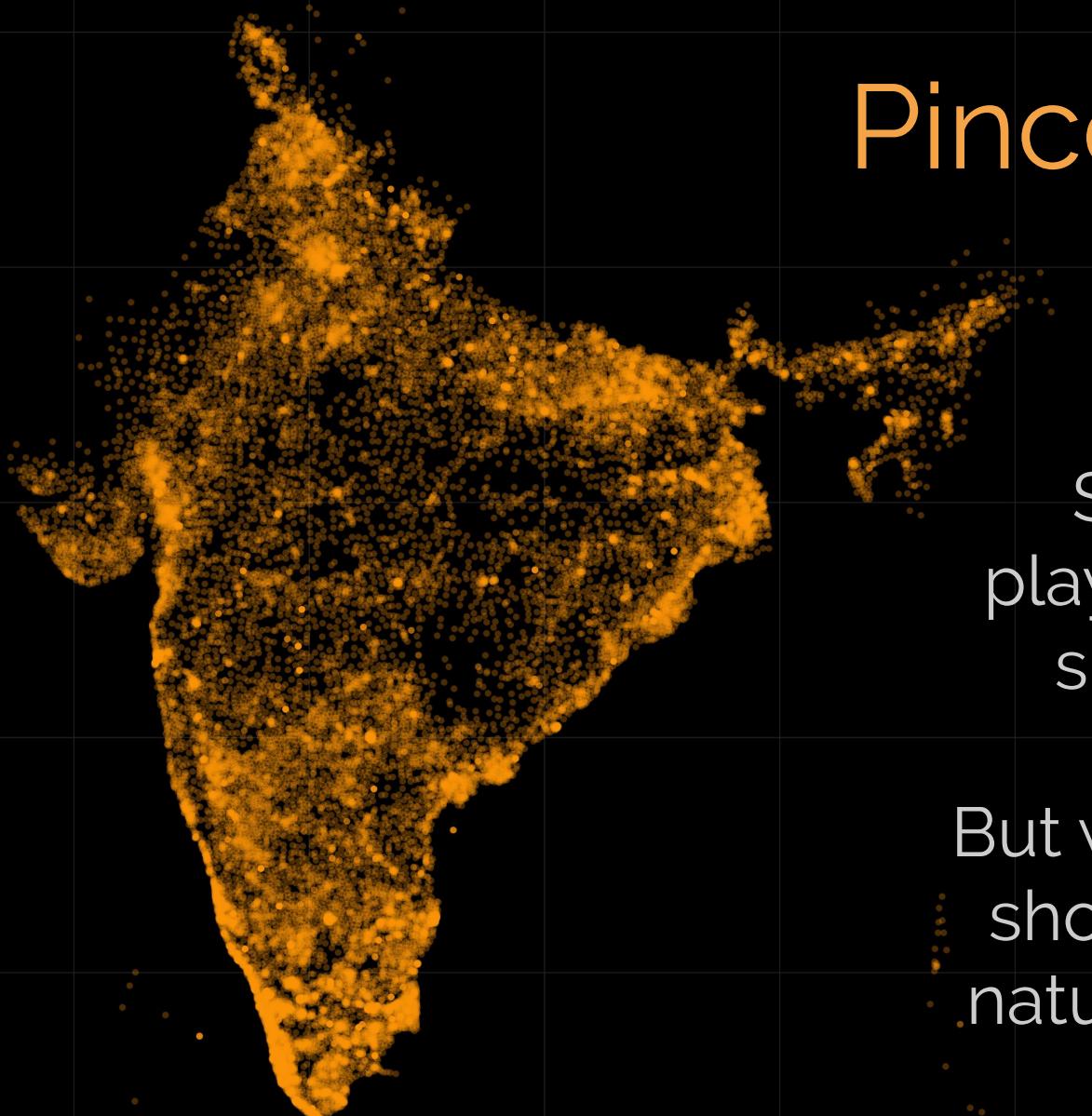
Pincode : 560076

Latitude : 12.8843049°

Longitude: 77.5967384°

Place : Bannerghatta

# Pincode Map



Scatter plot,  
play with alpha to  
show density

But what if I want to  
show geographic  
nature of pincode?

# Pincode+ Map

Visualisation of large  
data is iterative!

Refine Data  
(Filter, Transform)

## Acquire Data

Area	Sales
North	5
East	25
West	15
South	20
Central	10

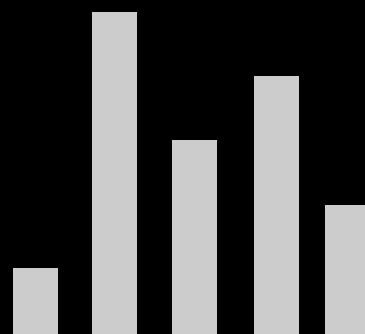
## Parse Variables

x	y
1	5
2	25
3	15
4	20
5	10

## Encode Shape & Select Scales

x	y
20	5
60	25
100	15
140	20
180	10

## Render with Coordinates



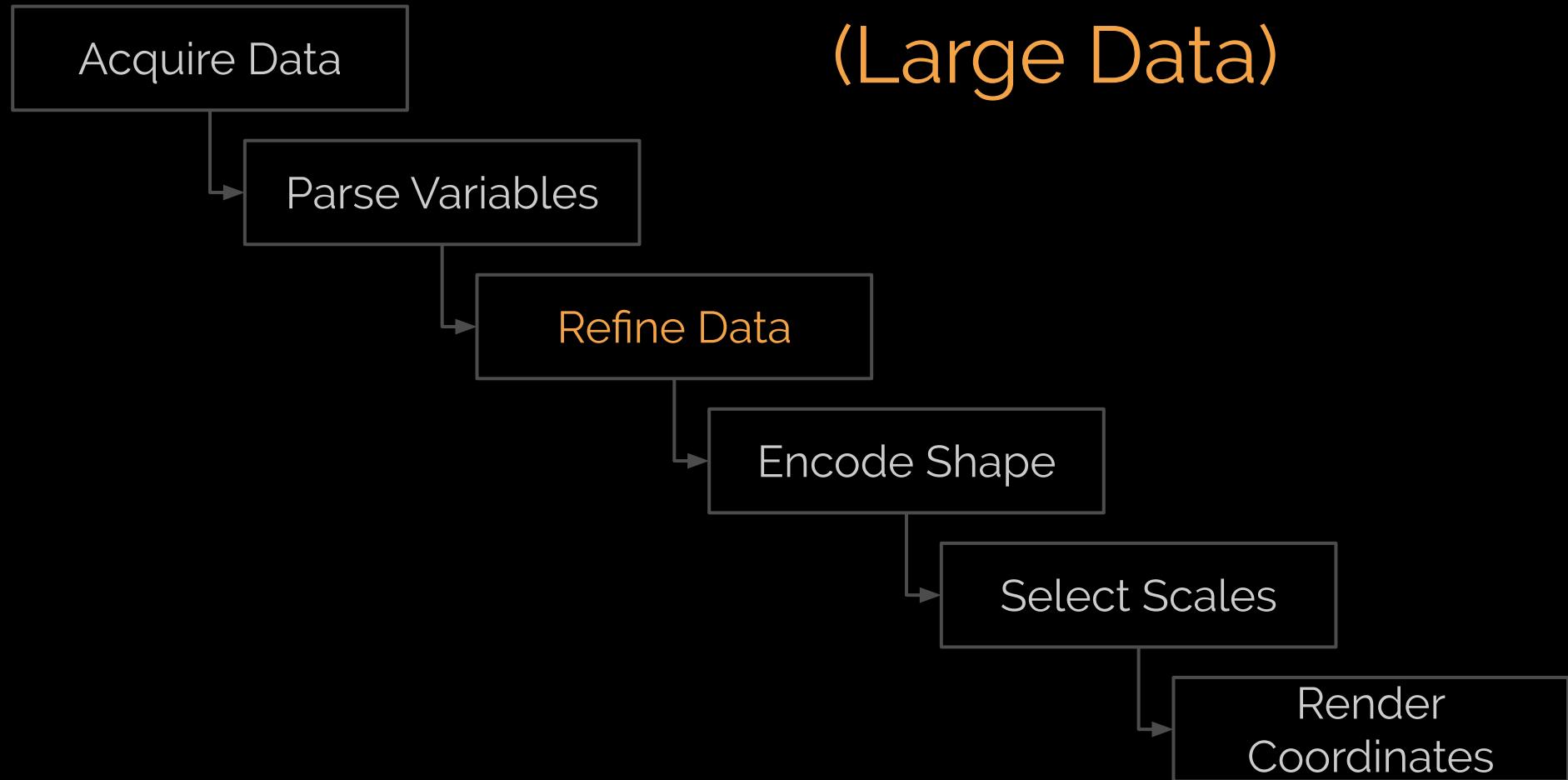
x (C) = Area  
y (Q) = Sales

x - position, y - bar scale - 200 x 200

cartesian

# Data Viz Process

## (Large Data)



# Diamonds dataset



50K+ observations of 10 dimensions

# Diamonds dataset



50K+ observations of 10 dimensions

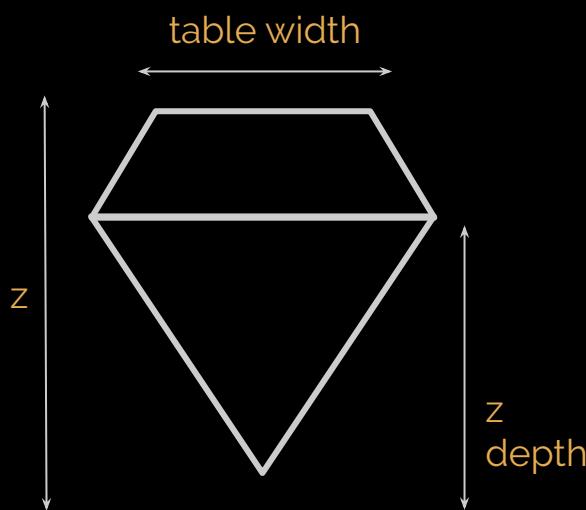
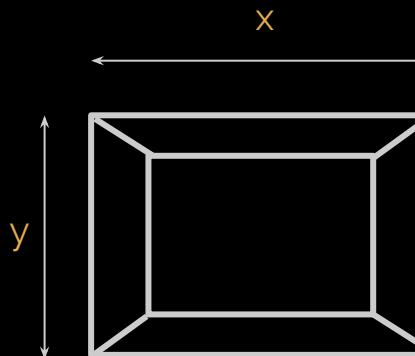
Price of  
diamonds is  
related to  
the 4C's

price	in US\$
carat	weight ( $\frac{1}{5}$ of a gram)
cut	5 levels [Fair to ideal]
colour	7 levels [J to D]
clarity	8 levels [I1 to IF]

# Diamonds dataset



50K+ observations of 10 dimensions



x	length mm
y	width mm
z	height mm
depth	z depth %
table	table width %



# Diamonds dataset

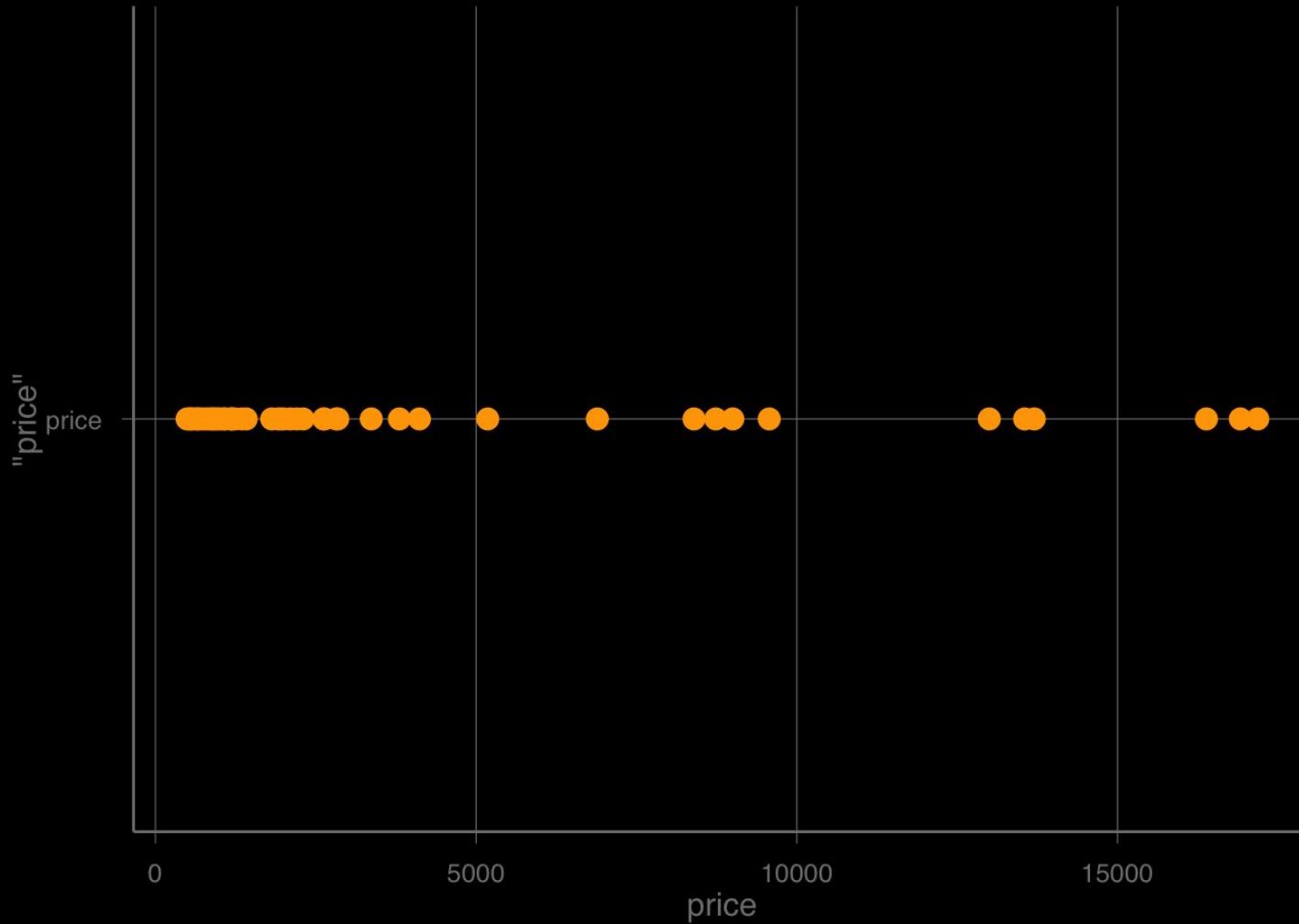
50K+ observations of 10 dimensions

price	carat	cut	color	clarity	x	y	z	depth	table
326	0.23	Ideal	E	SI2	3.95	3.98	2.43	61.5	55
326	0.21	Premium	E	SI1	3.89	3.84	2.31	59.8	61
327	0.23	Good	E	VS1	4.05	4.07	2.31	56.9	65
334	0.29	Premium	I	VS2	4.2	4.23	2.63	62.4	58
335	0.31	Good	J	SI2	4.34	4.35	2.75	63.3	58
336	0.24	Very Good	J	VVS2	3.94	3.96	2.48	62.8	57

**1d**

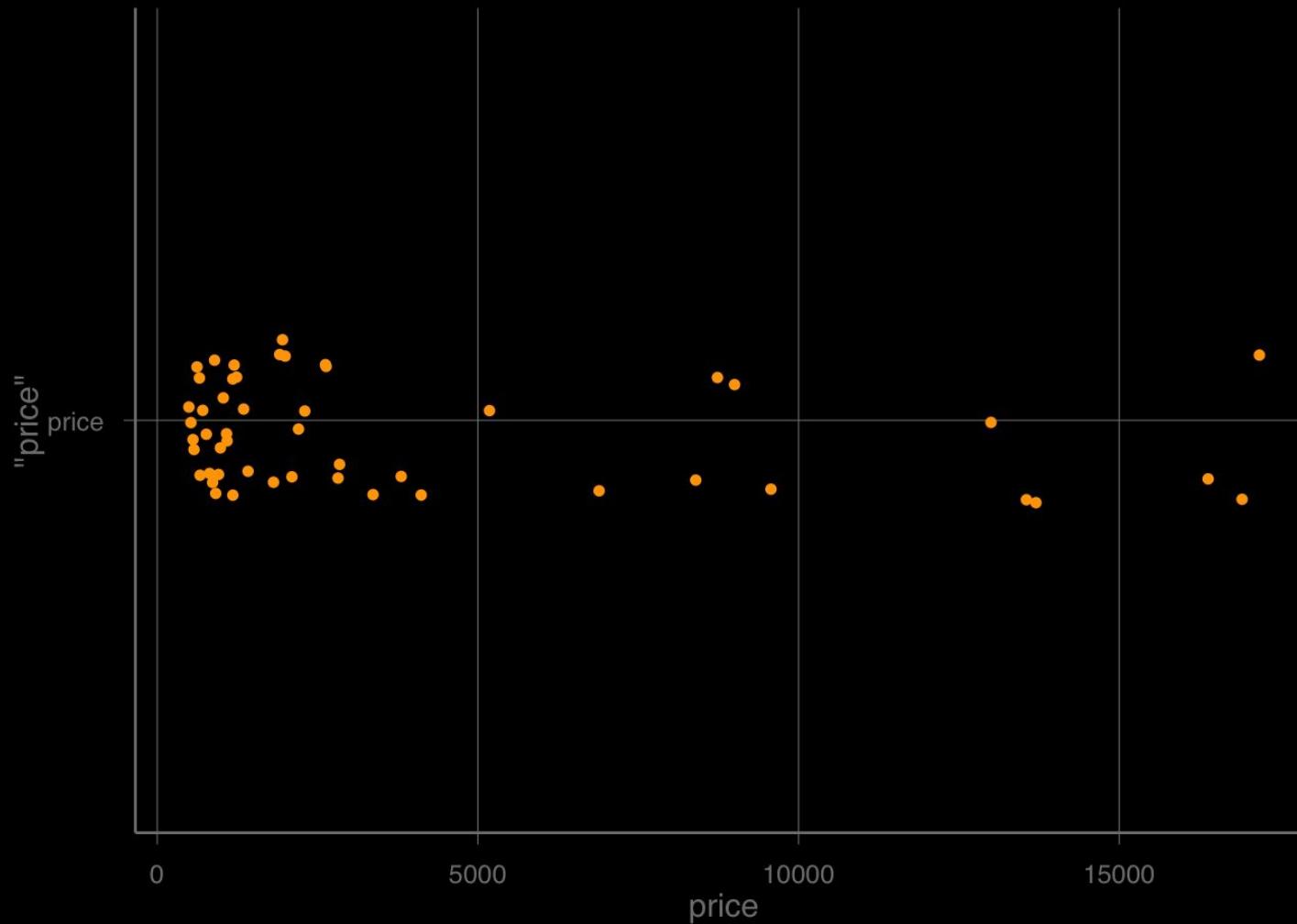


# Strip Plot



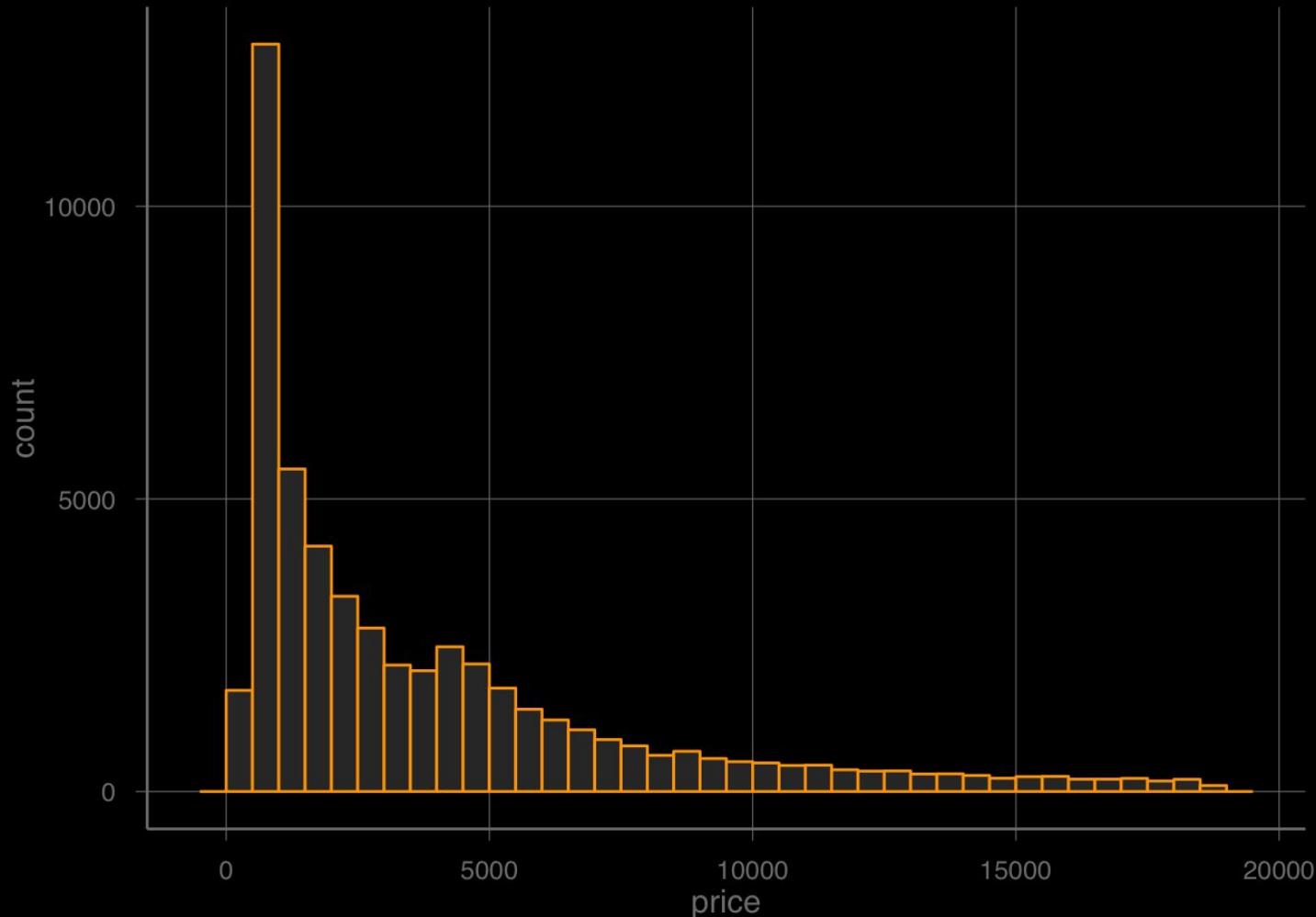
Data : 1 x quantitative  
Encoding : position

# Strip Plot - Jitter



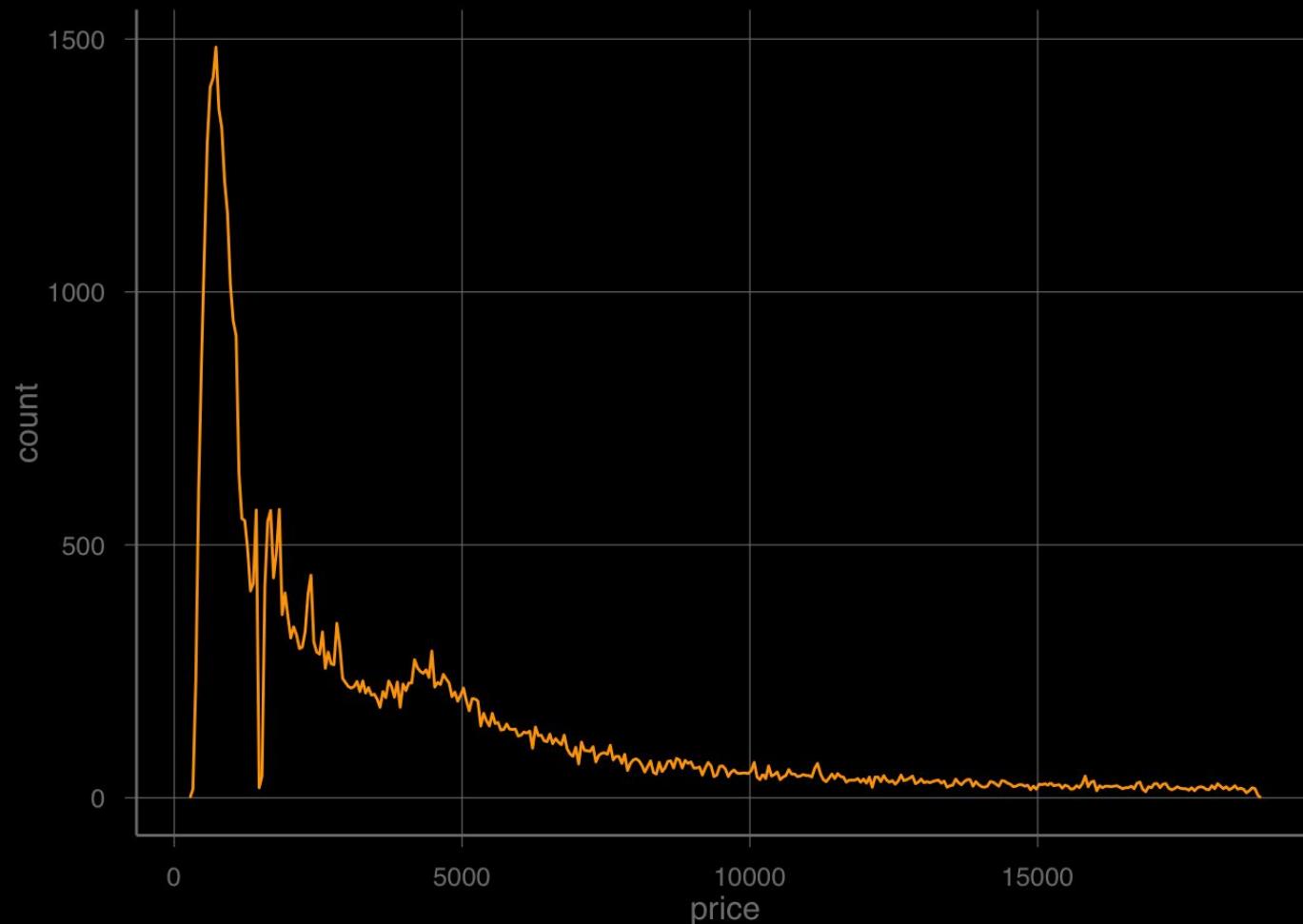
Data : 1 x quantitative  
Encoding : position

# Histogram



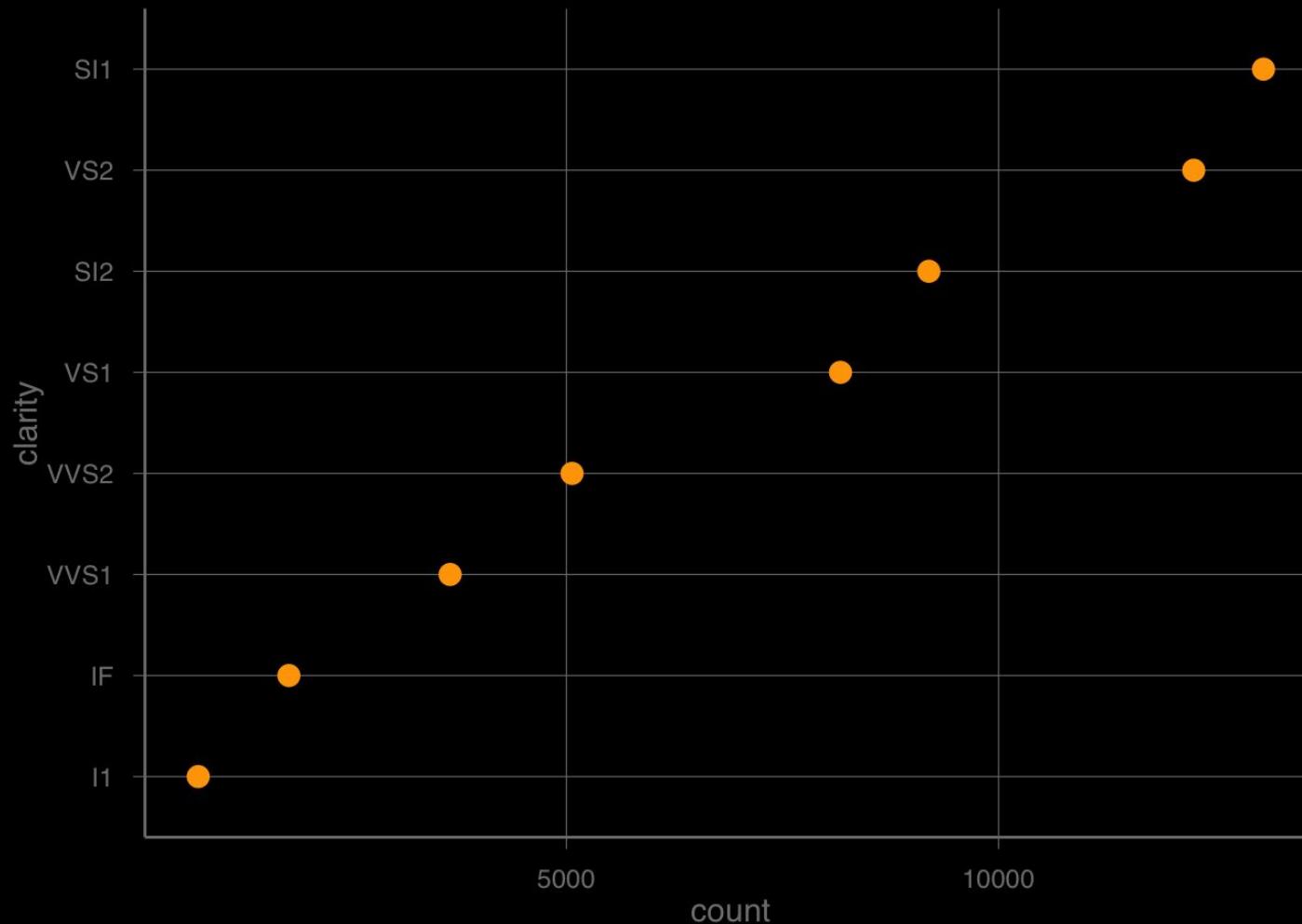
Data : 1 x quantitative  
Encoding : length

# Freq Polygon



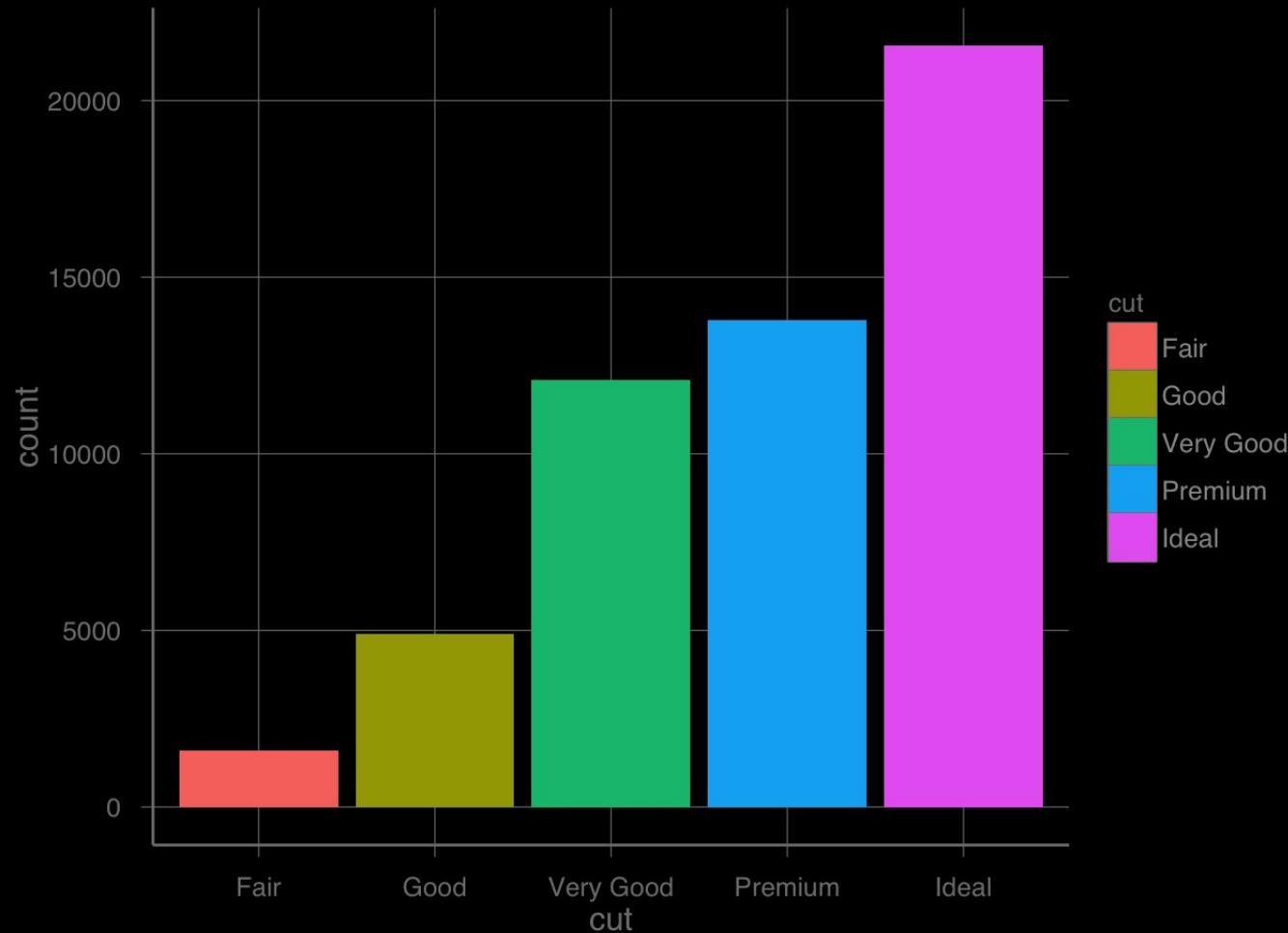
Data : 1 x quantitative  
Encoding : line

# Dot-Plot



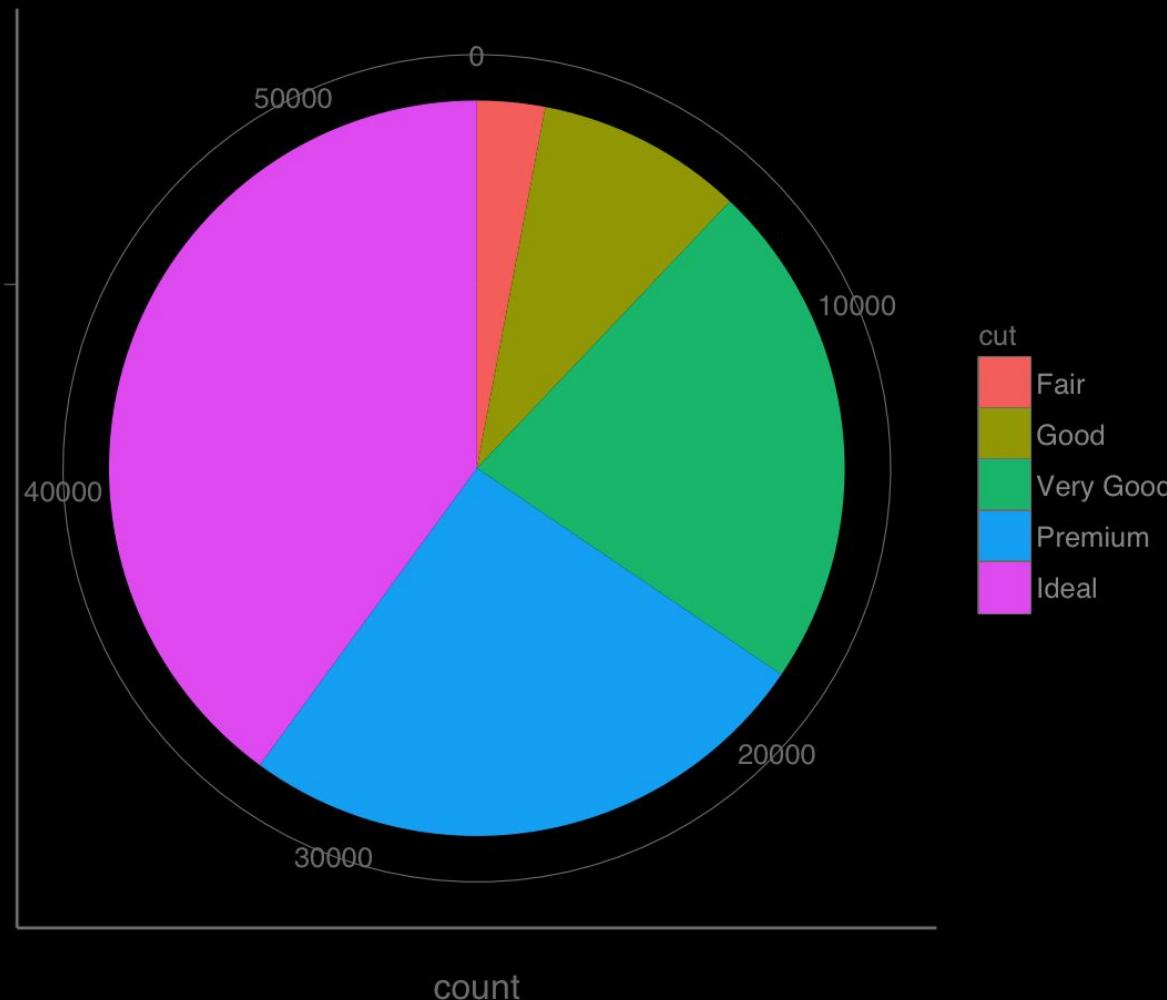
Data : 1 x categorical  
Encoding : position

# Bar (Column) Chart



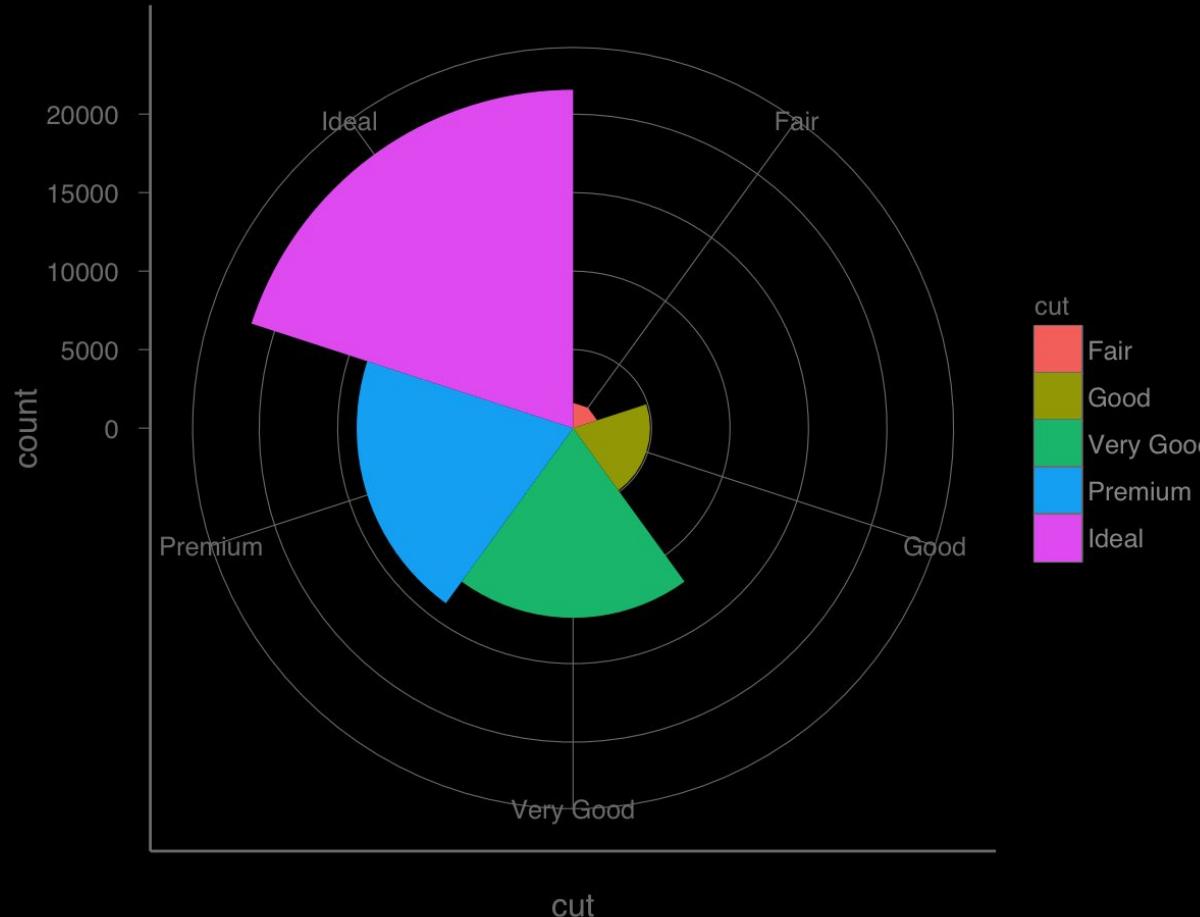
Data : 1 x categorical  
Encoding : length, (color)

# Pie Chart



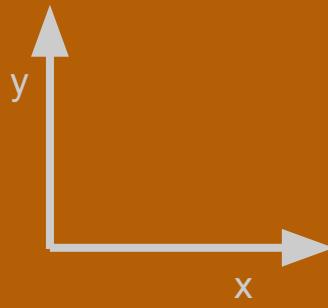
Data : 1 x categorical  
Encoding : angle, (color)

# Coxcomb Chart

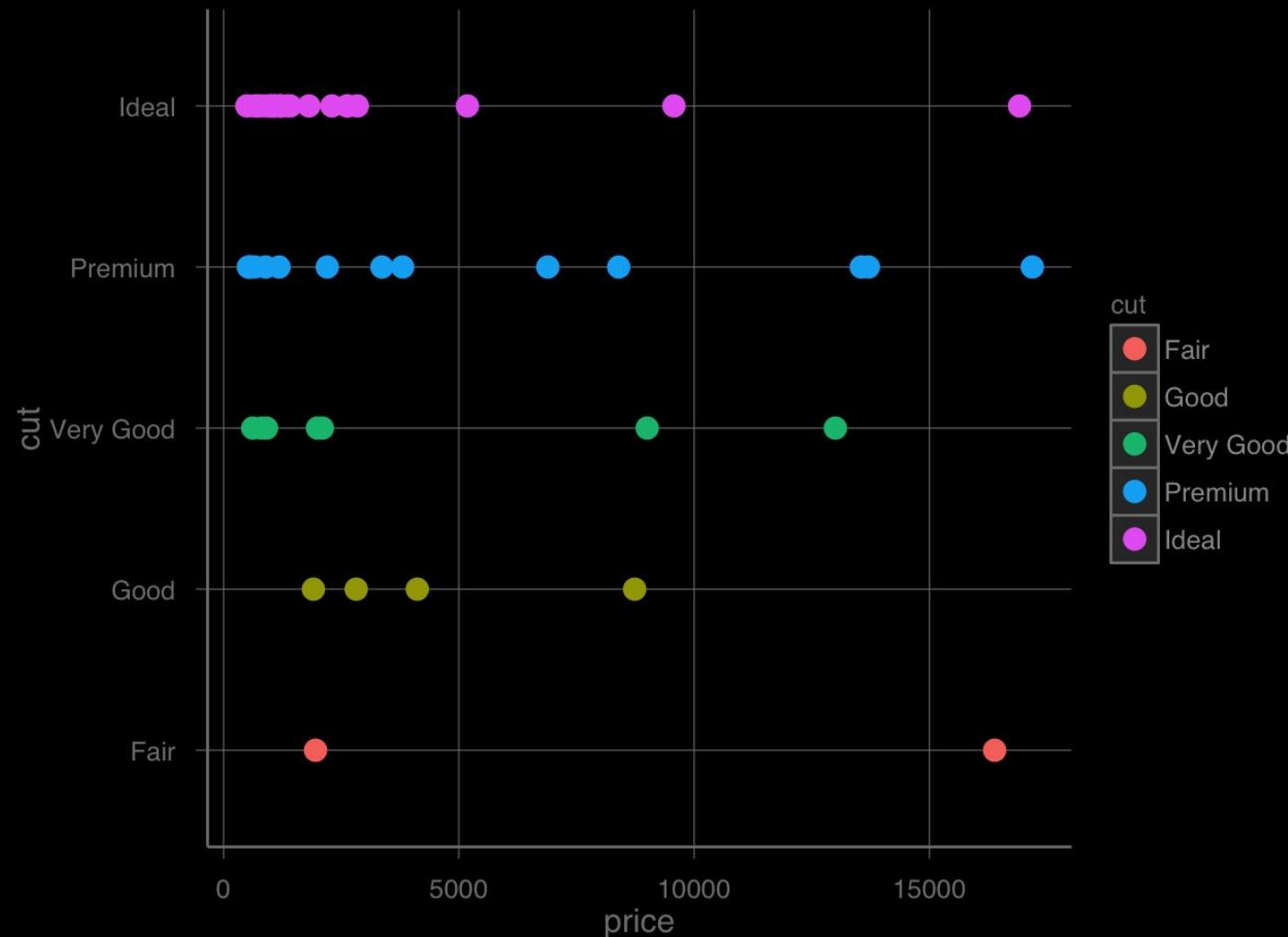


Data : 1 x categorical  
Encoding : length, color, angle

2d

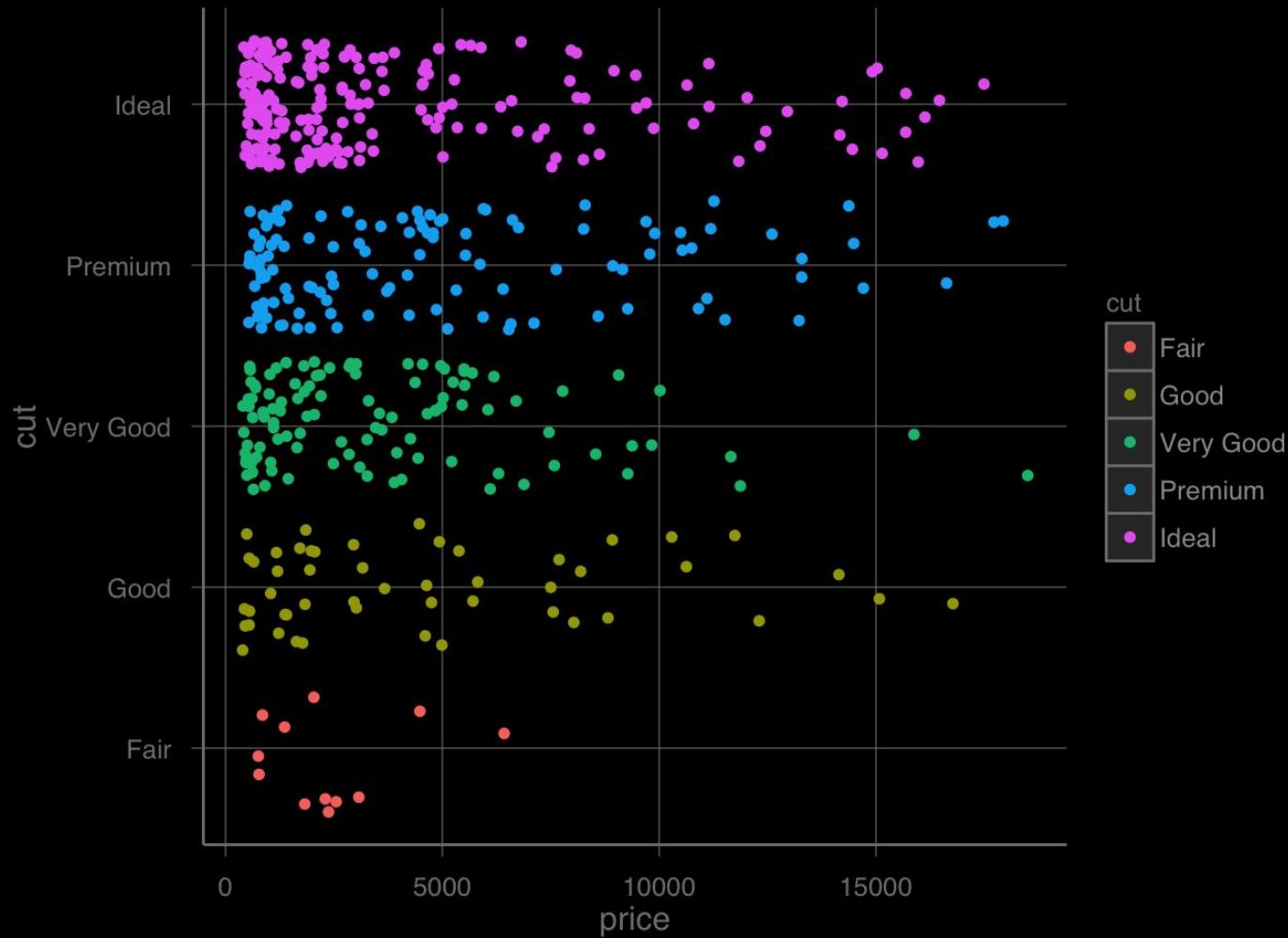


# Strip Plot - Multiple



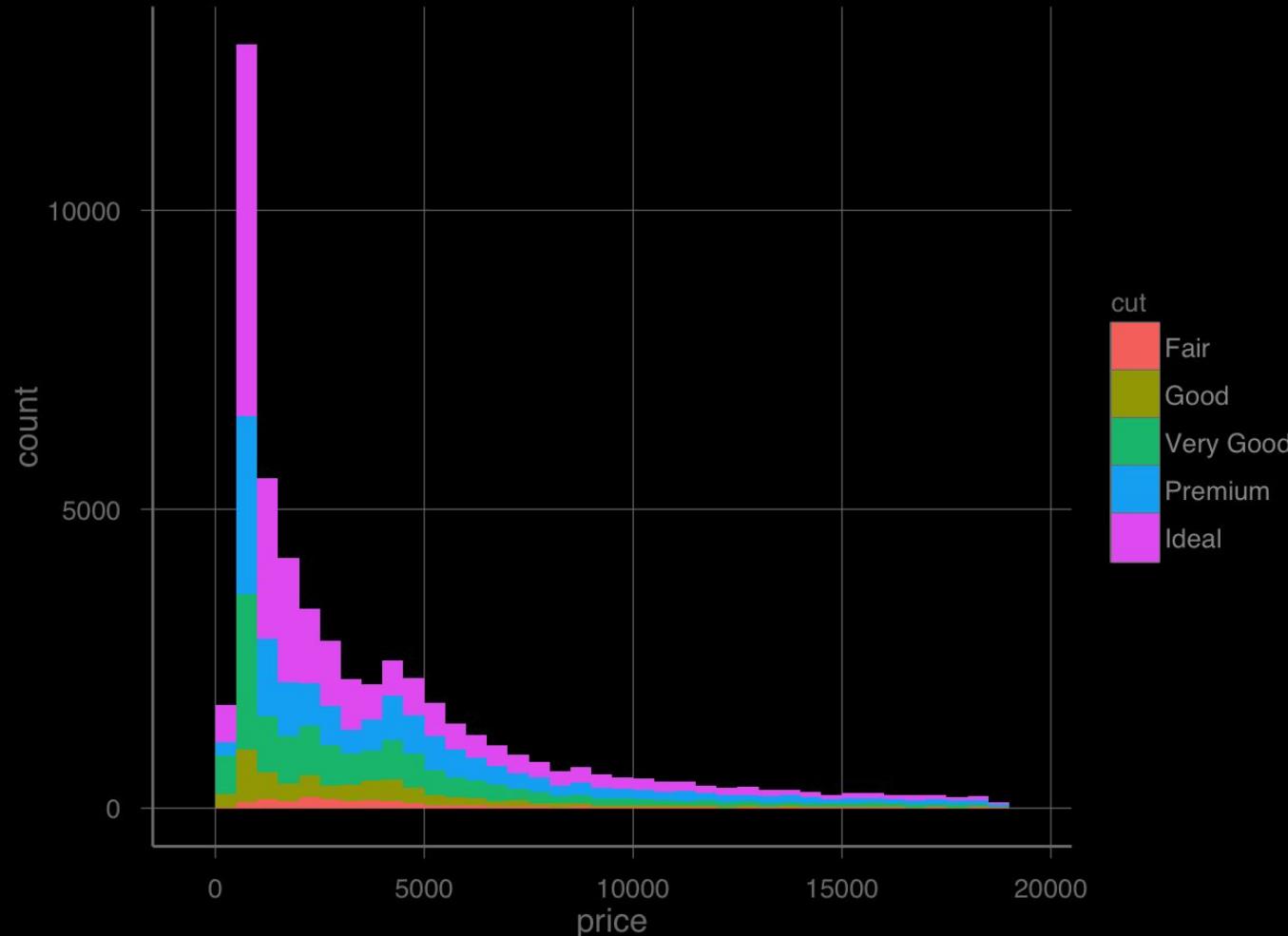
Data : 1 x quantitative, 1 x categorical  
Encoding : position, (color)

# Strip Plot - Multiple (J)



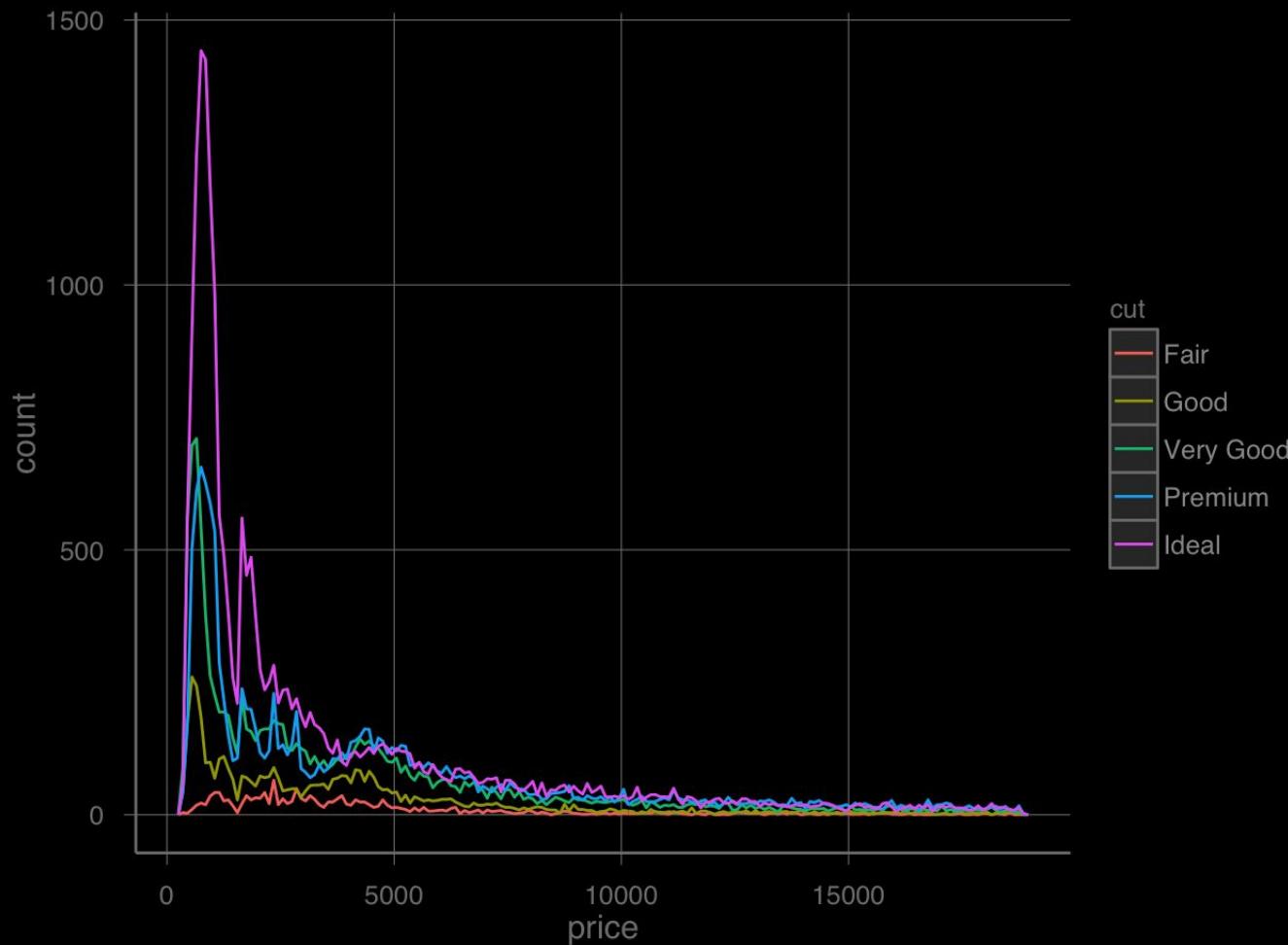
Data : 1 x quantitative, 1 x categorical  
Encoding : position, (color)

# Histogram - Multiple



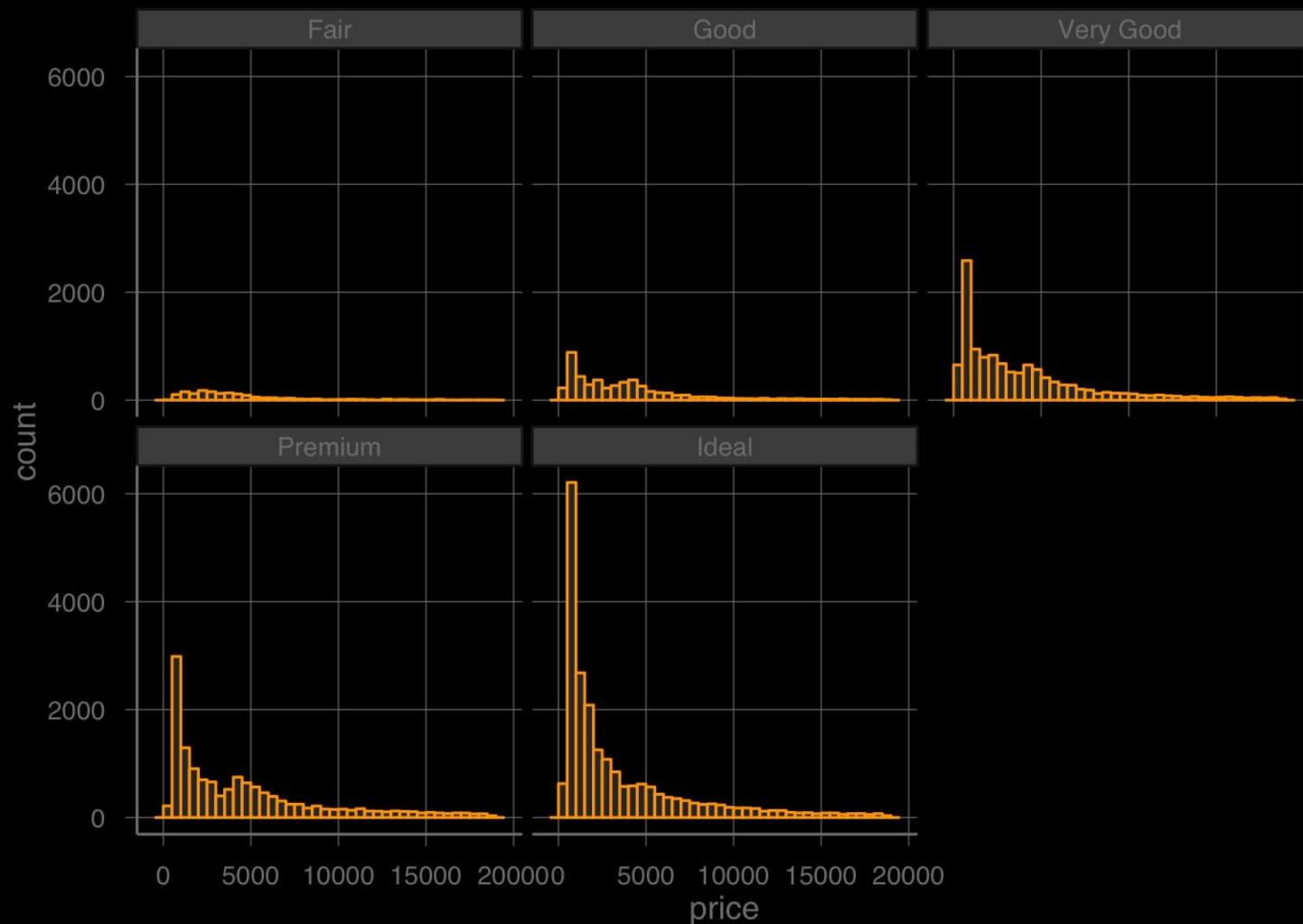
Data : 1 x quantitative, 1 x categorical  
Encoding : length, color

# Freq Polygon-Multiple



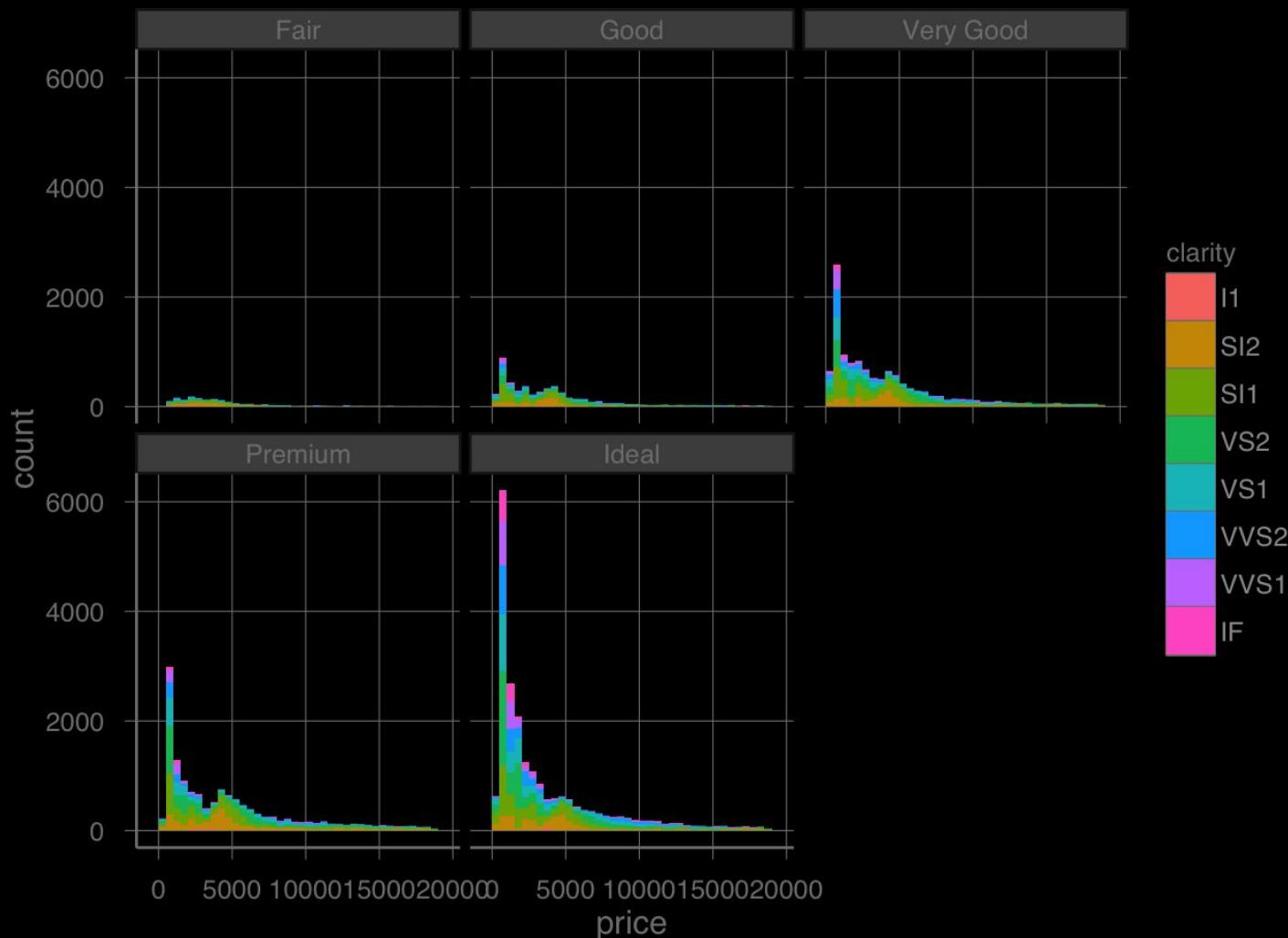
Data : 1 x quantitative, 1 x categorical  
Encoding : line, color

# Small Multiples - One



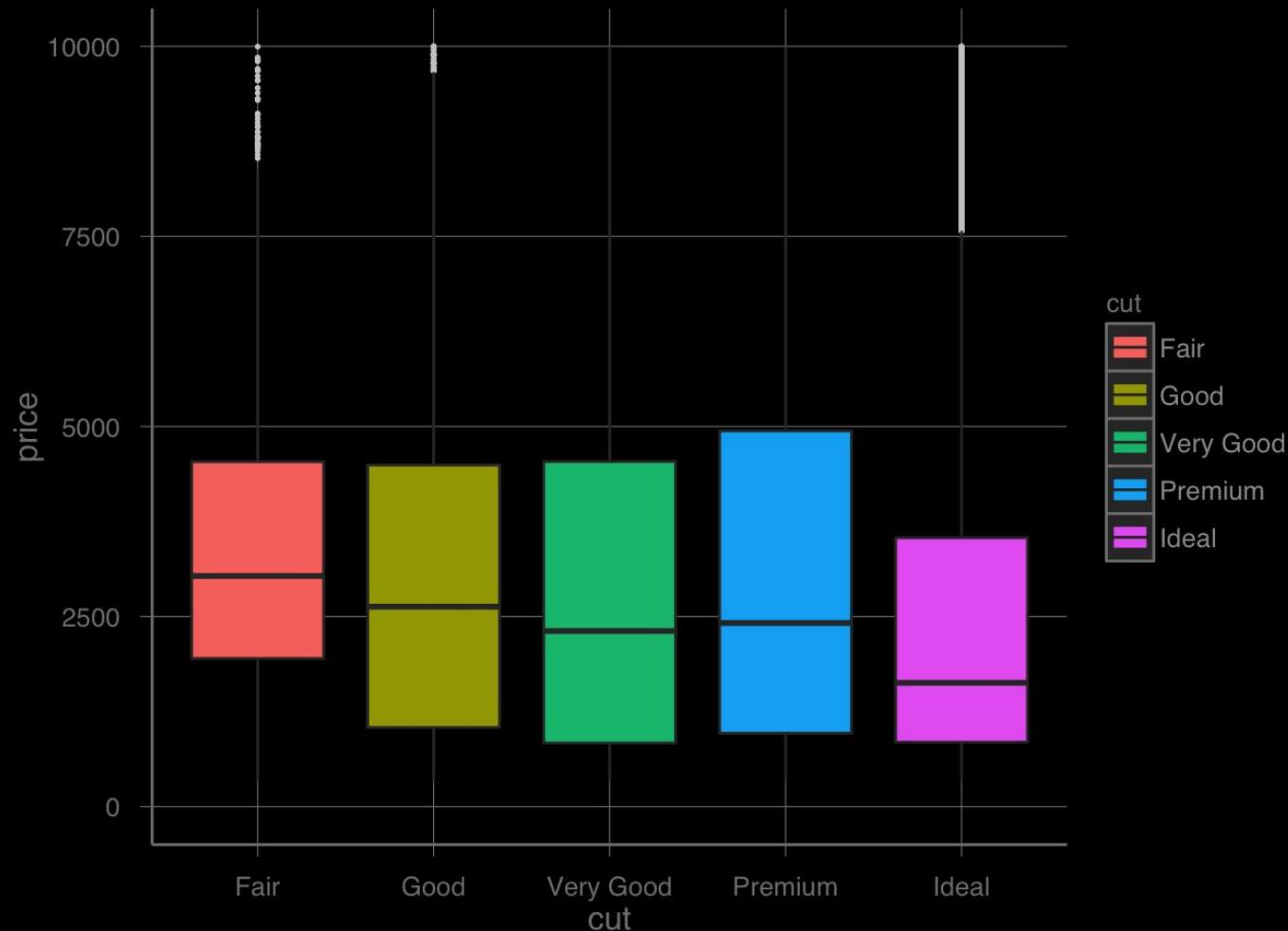
Data : 1 x quantitative, 1 x categorical  
Encoding : length, color, arrangement

# Small Multiples - Two



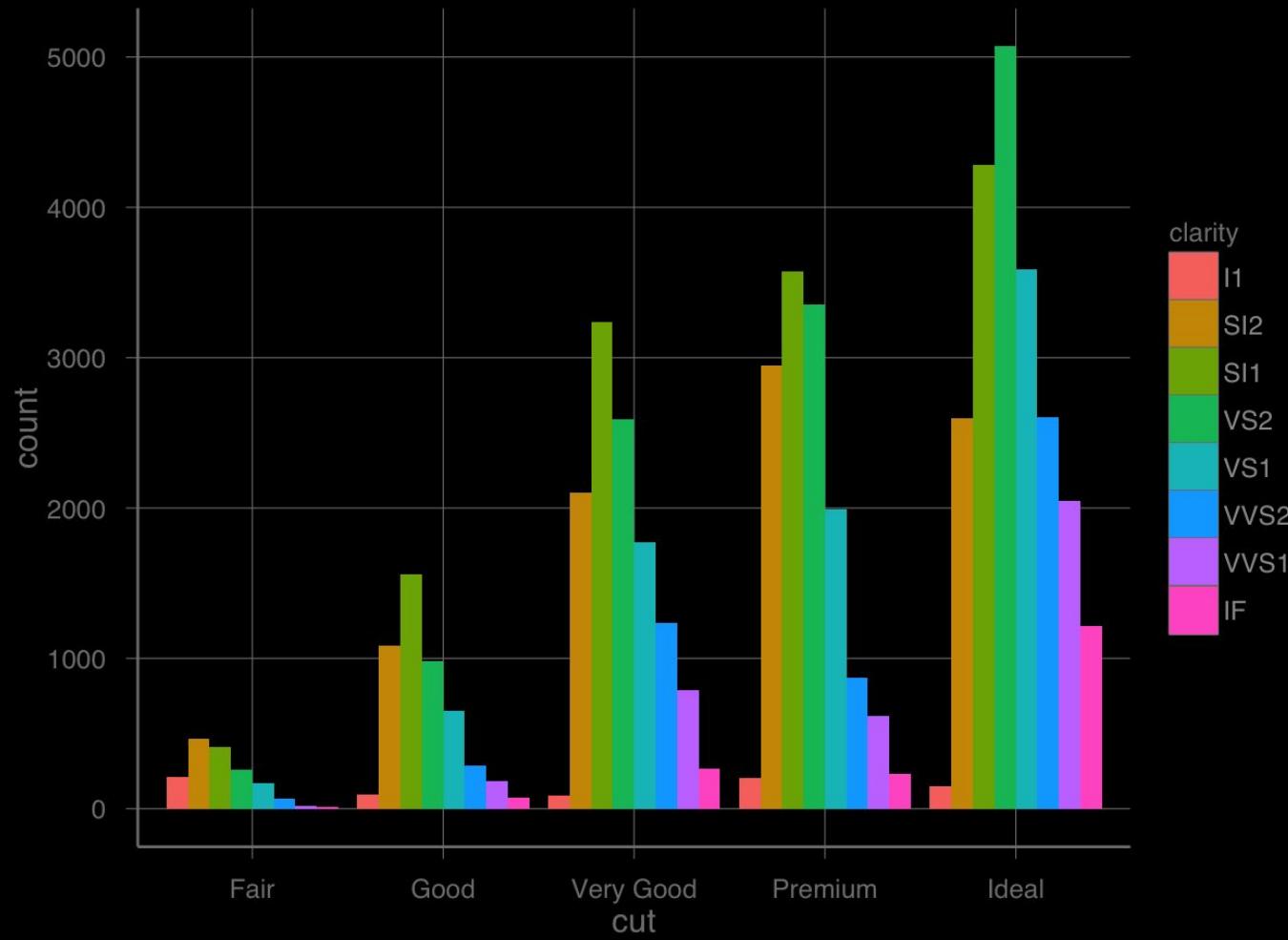
Data : 1 x quantitative, 2 x categorical  
Encoding : length, color, arrangement

# Box Plot - Multiple



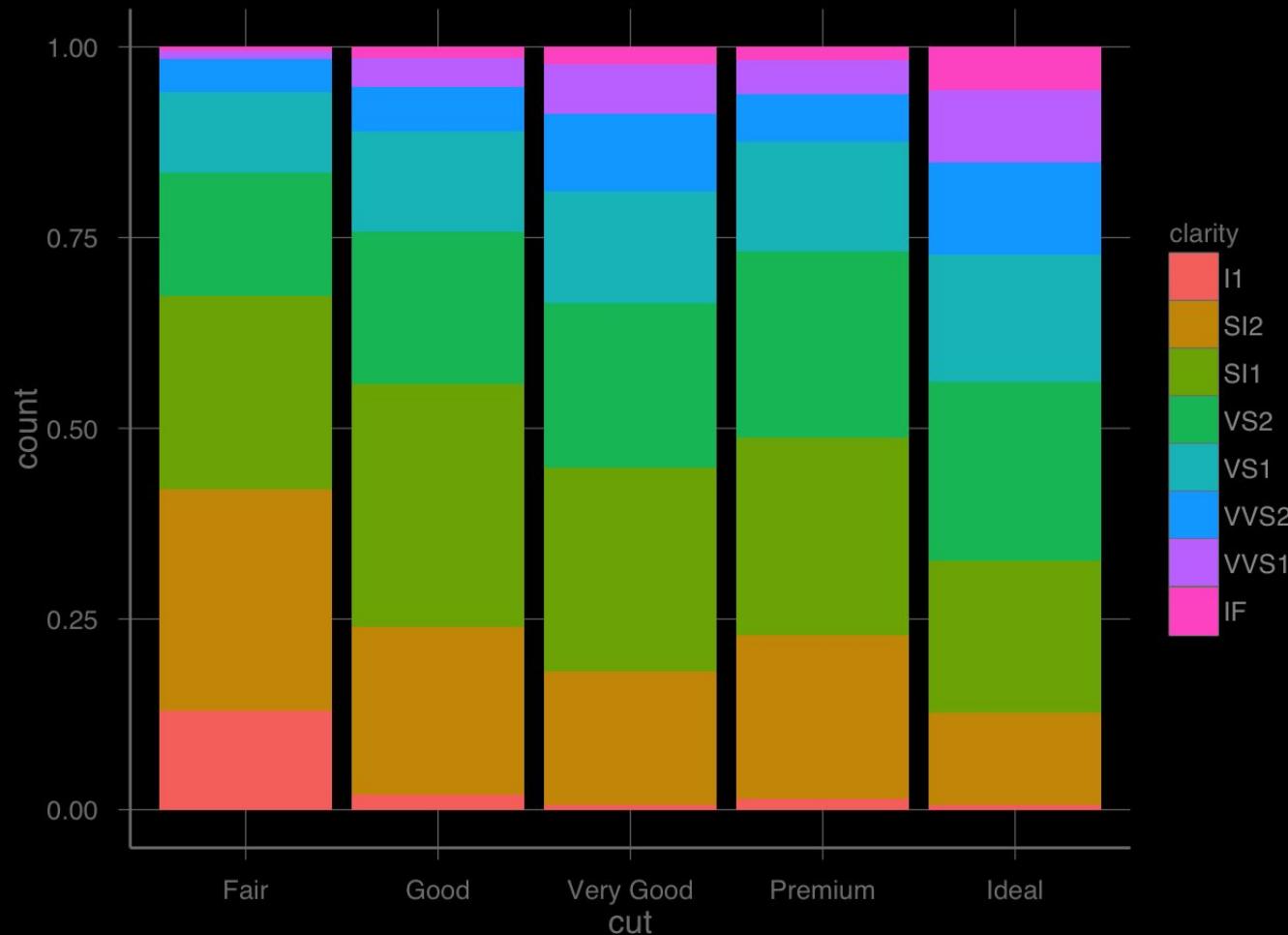
Data : 1 x quantitative, 1 x categorical  
Encoding : position, length, color

# Bar (Column) Chart



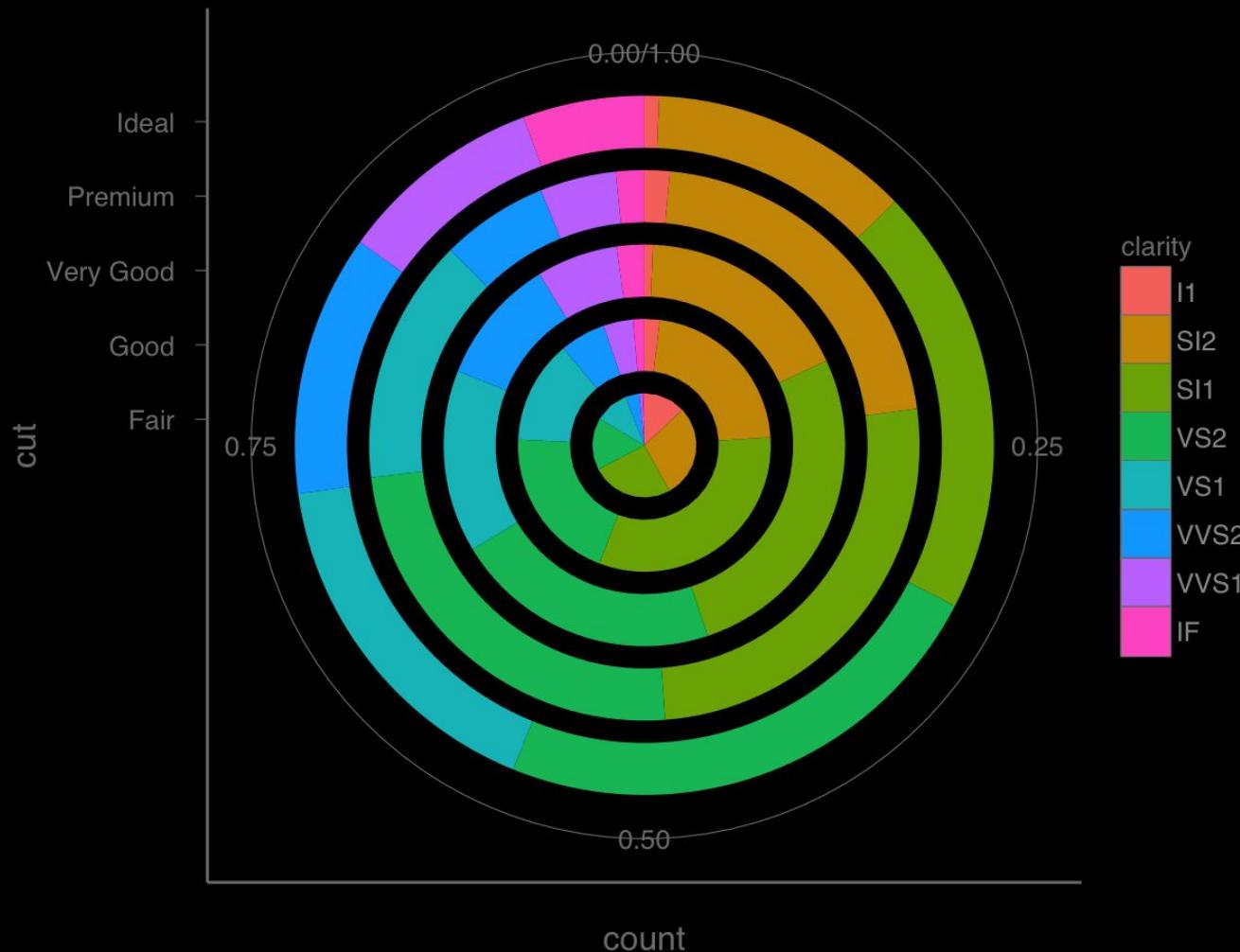
Data : 1 x quantitative, 2 x categorical  
Encoding : length, color

# Stacked Bar - 100%



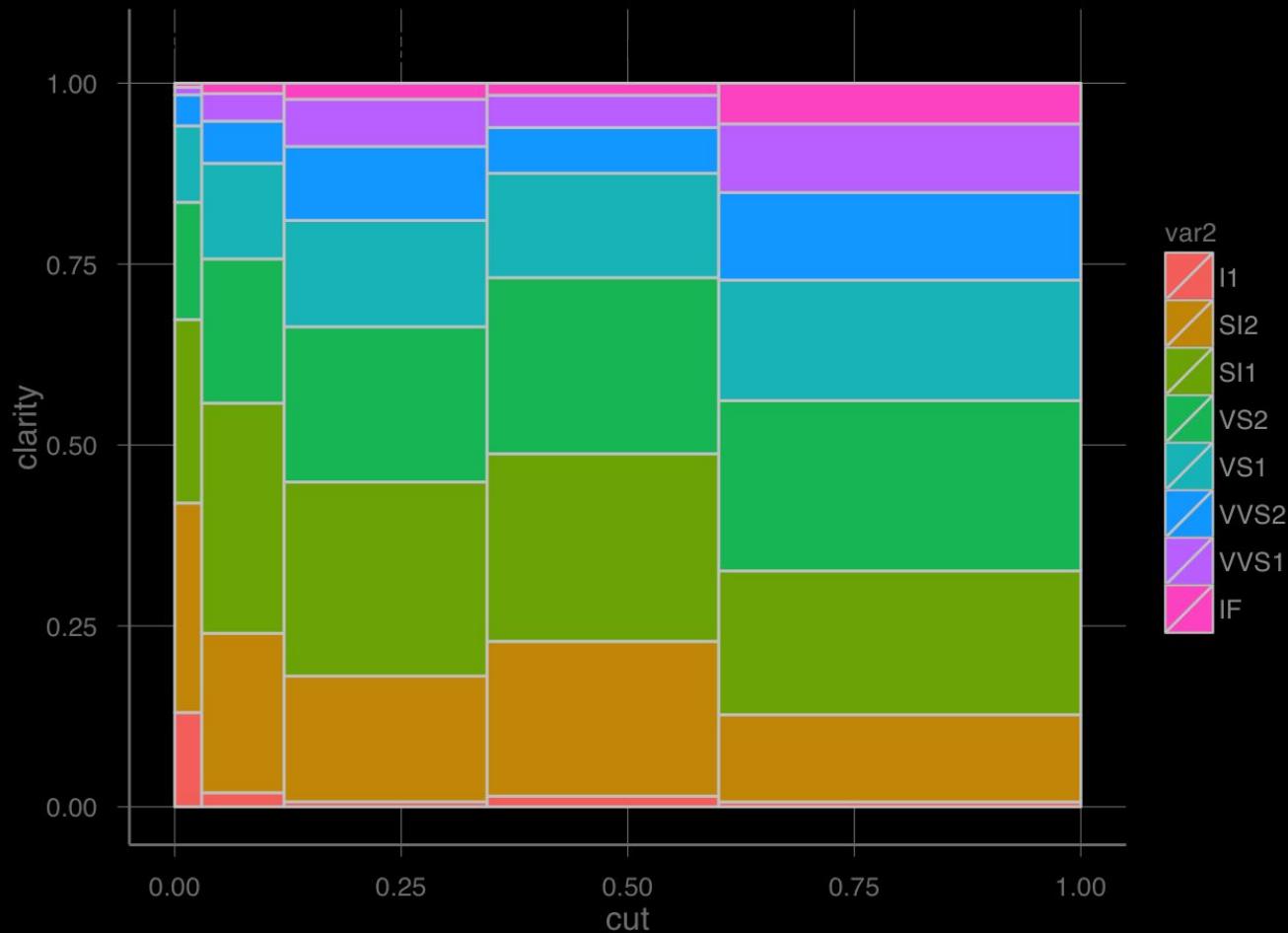
Data : 1 x quantitative, 2 x categorical  
Encoding : length, color, position

# Donuts



Data : 1 x quantitative, 2 x categorical  
Encoding : angle, color, length

# MariMekko or Mosaic



Data : 1 x quantitative, 2 x categorical  
Encoding : area, color

1d



2d

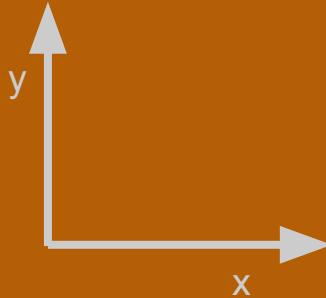


Chart Options

Points



Bars



Lines



Areas

1d Quantitative

1d Categorical

2d Quantitative +  
Categorical

2d Categorical +  
Categorical

2d Quantitative +  
Quantitative

## Chart Options

● Points

Bars

Lines

Areas

### 1d Quantitative



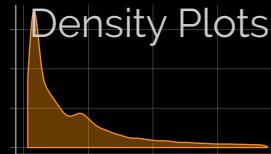
Strip Plot



Histogram



Freq Poly

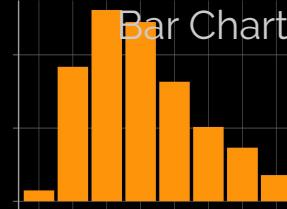


Density Plots

### 1d Categorical



Dot Plot

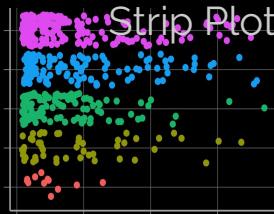


Bar Chart

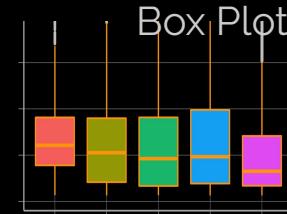
Avoid

Avoid

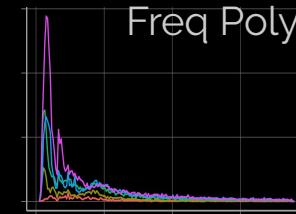
### 2d Quantitative + Categorical



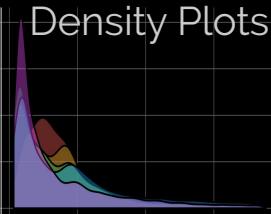
Strip Plot



Box Plot



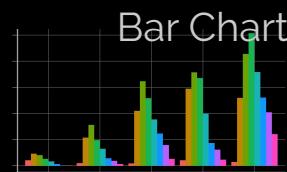
Freq Poly



Density Plots

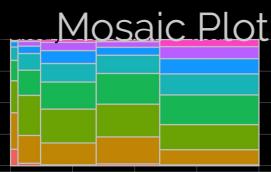
### 2d Categorical + Categorical

Avoid



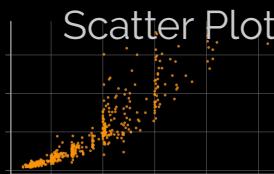
Bar Chart

Avoid



Mosaic Plot

### 2d Quantitative + Quantitative



Scatter Plot

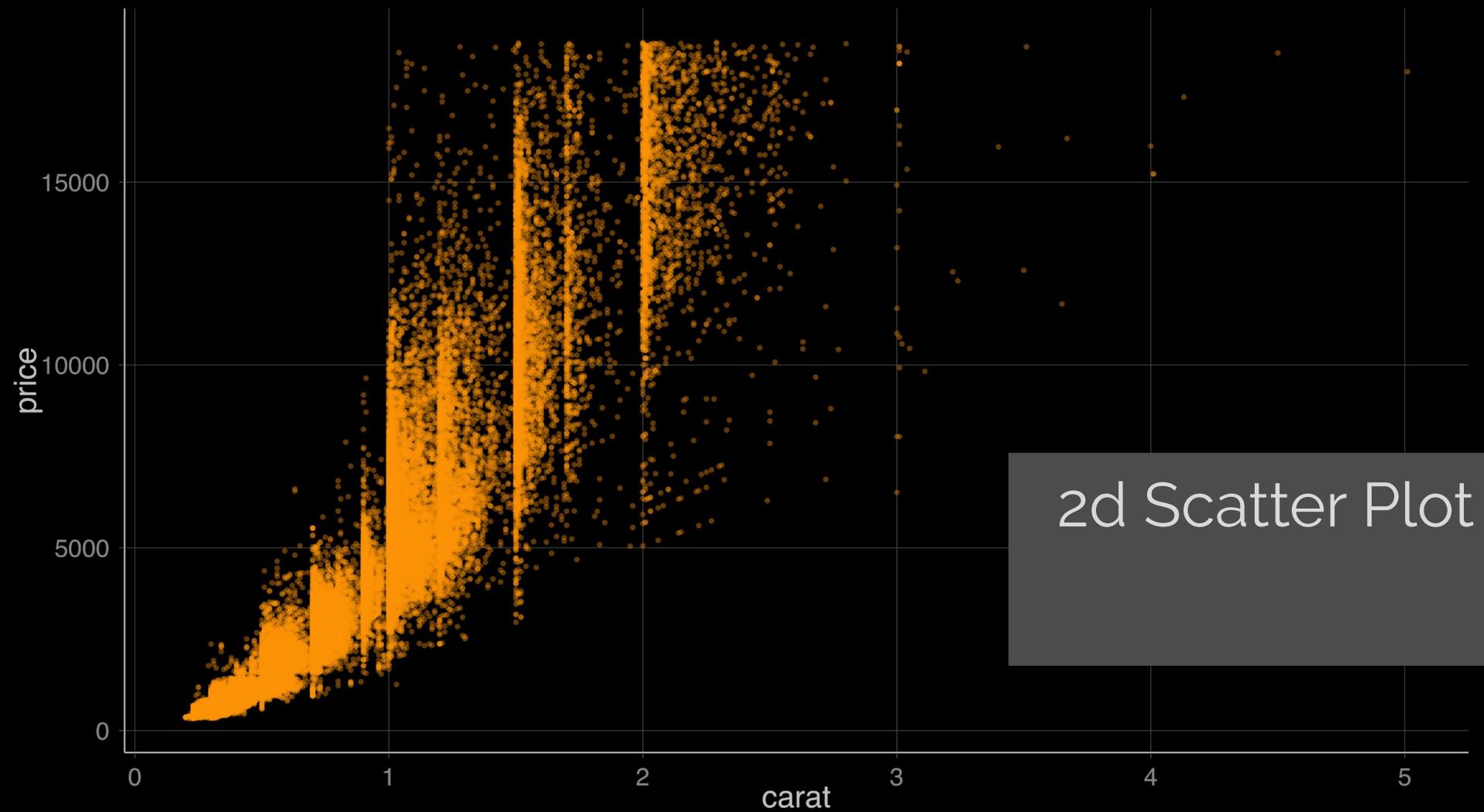


Table Lens

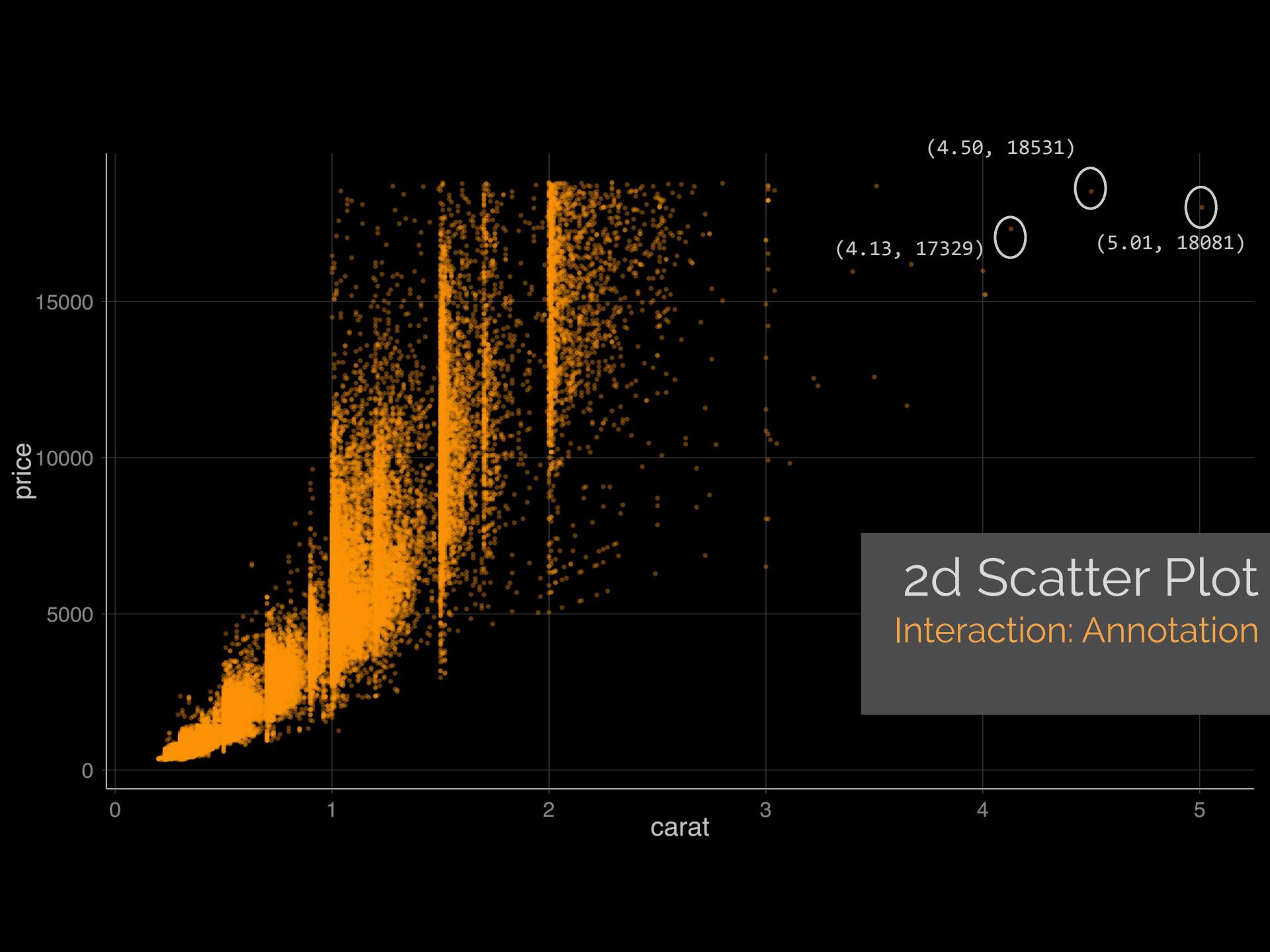


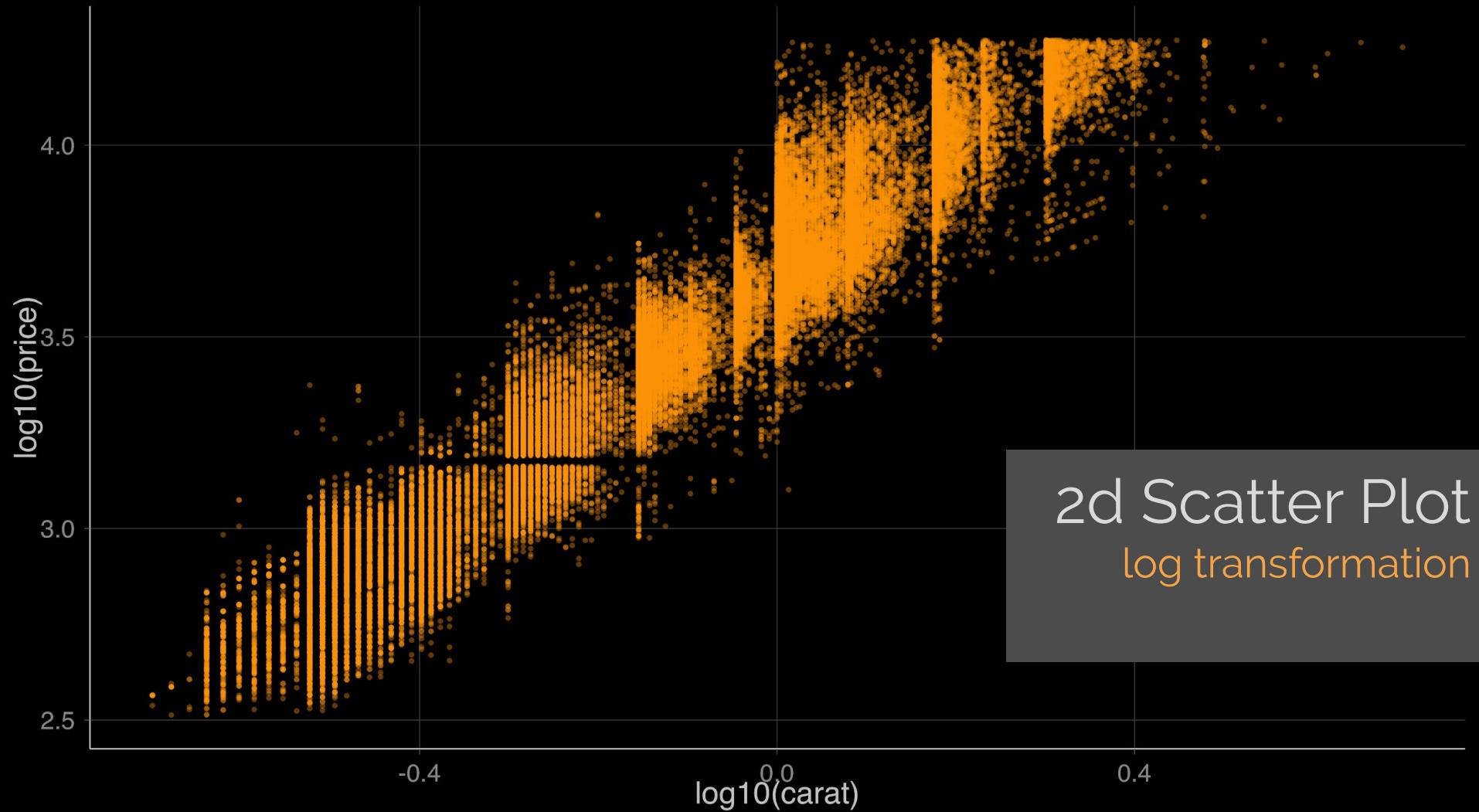
Slopegraph

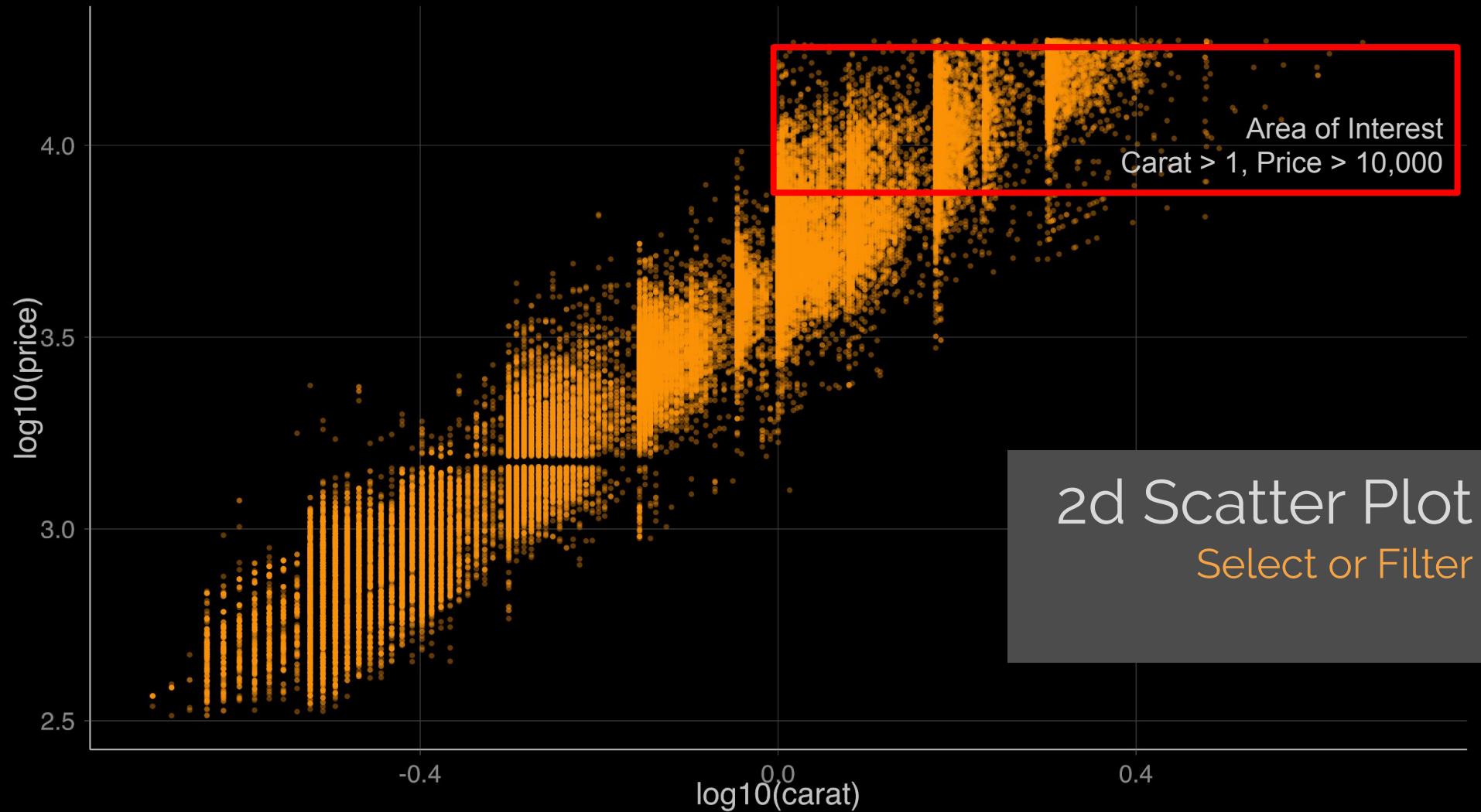
Avoid



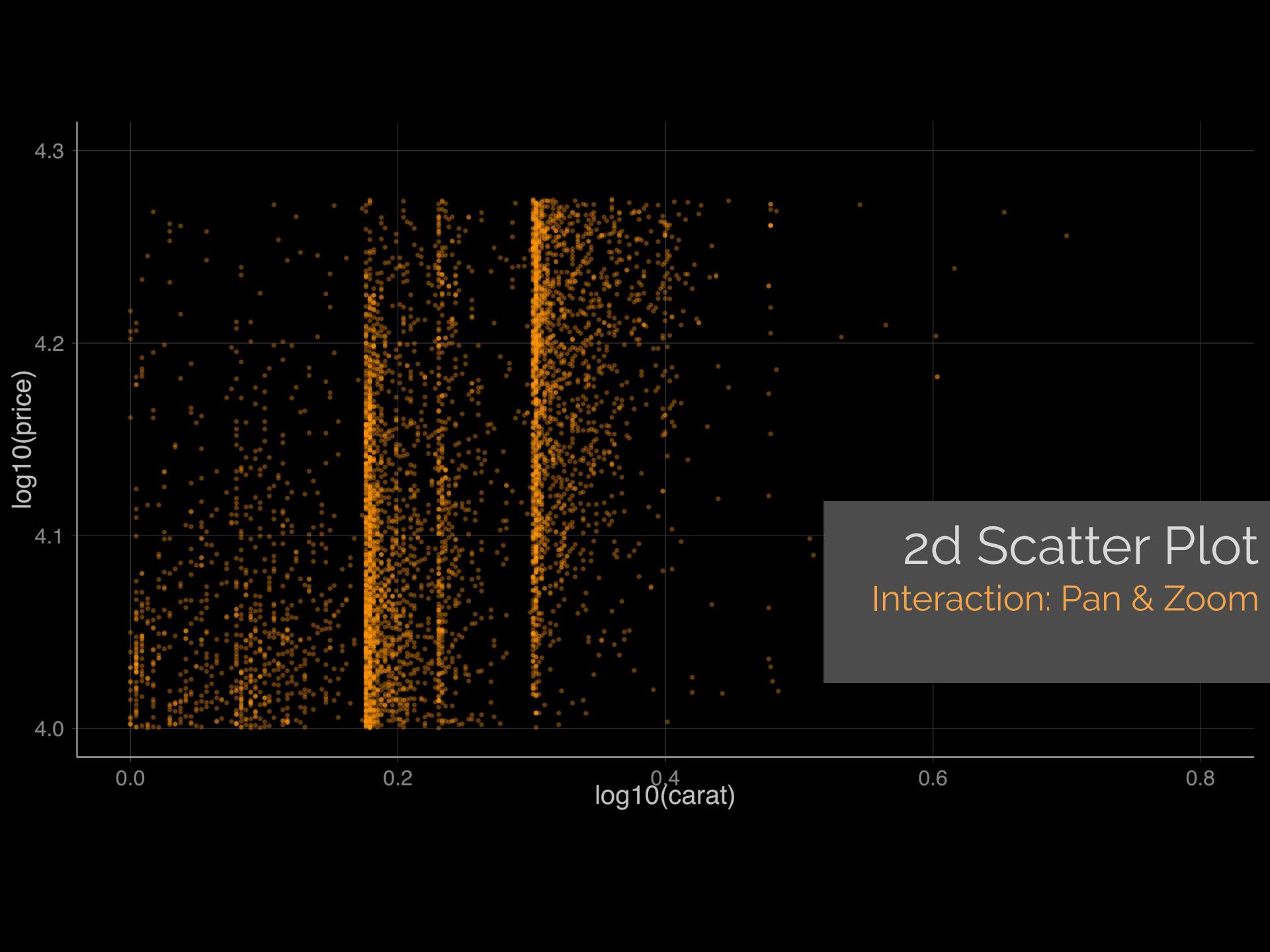
2d Scatter Plot



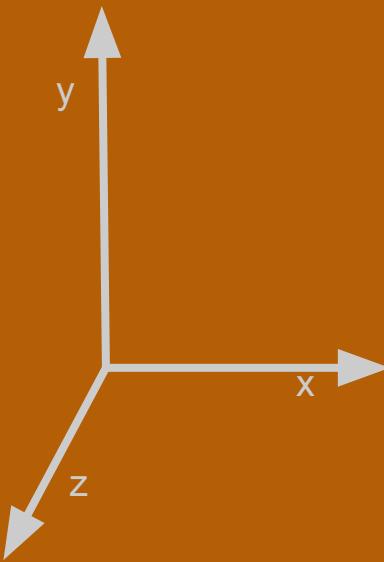




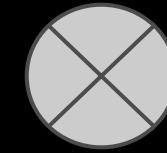
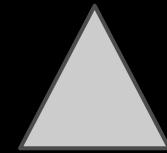
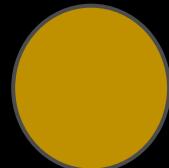
2d Scatter Plot  
Select or Filter



3d



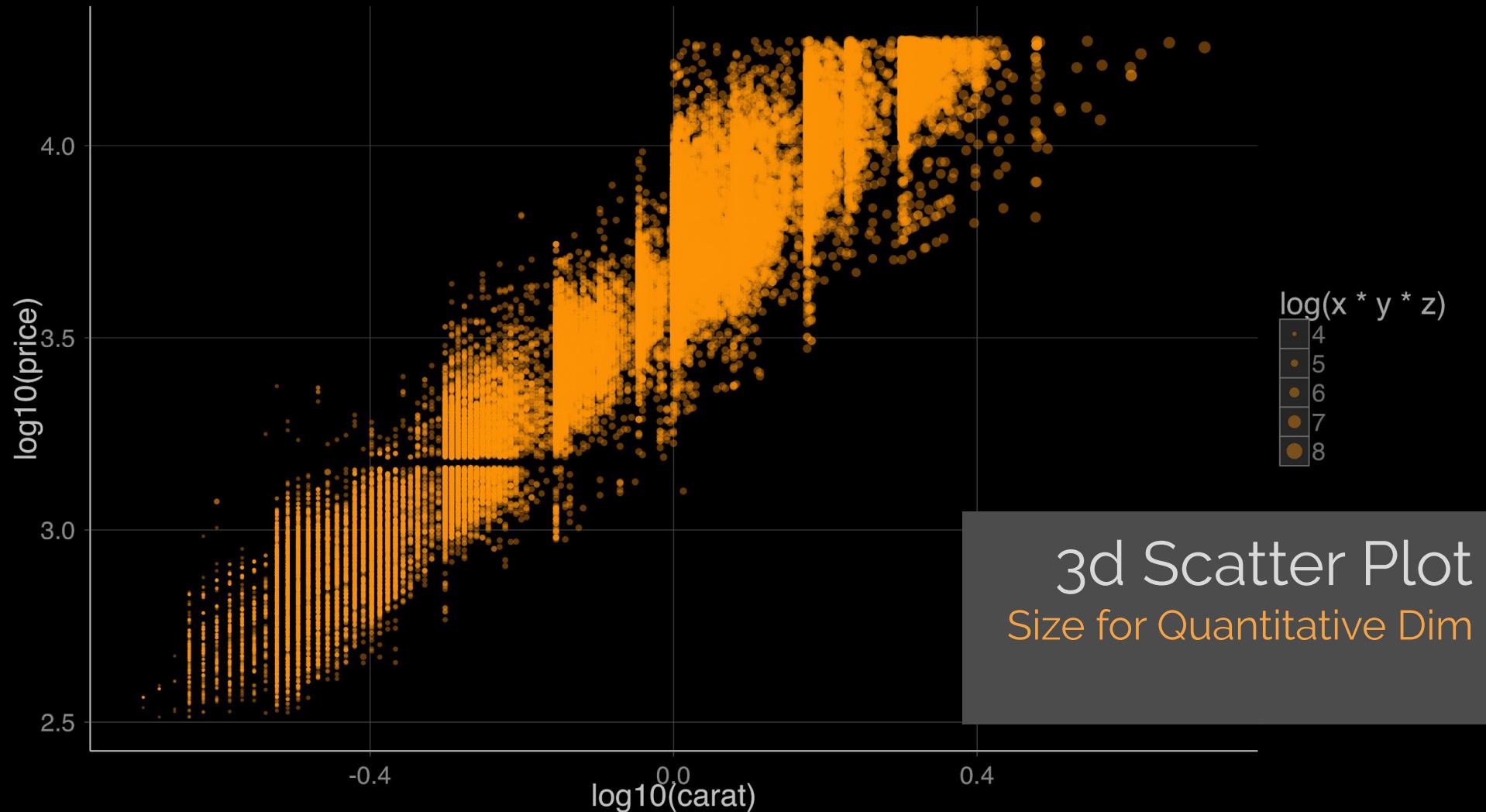
# Use shape for 3d

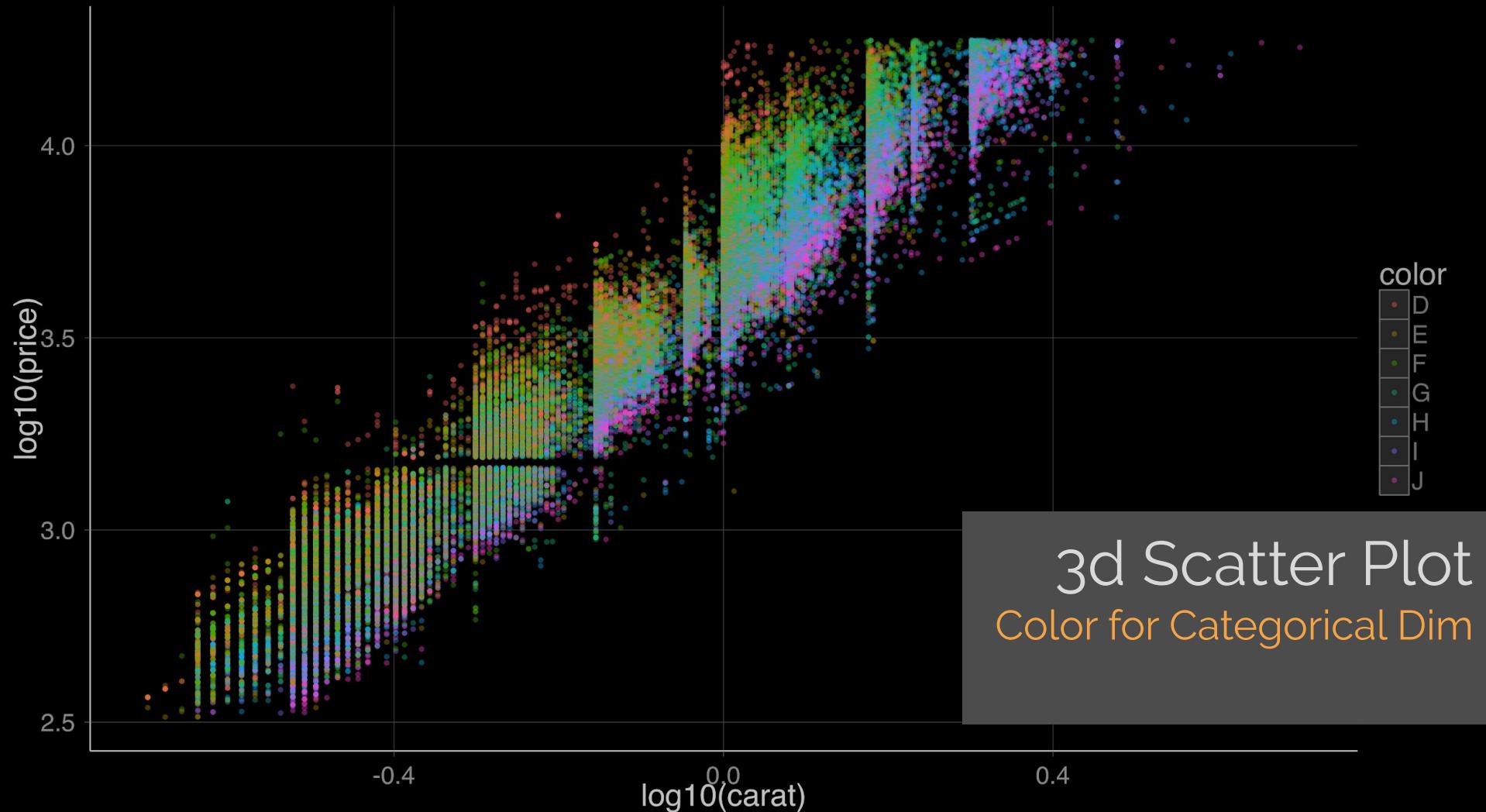


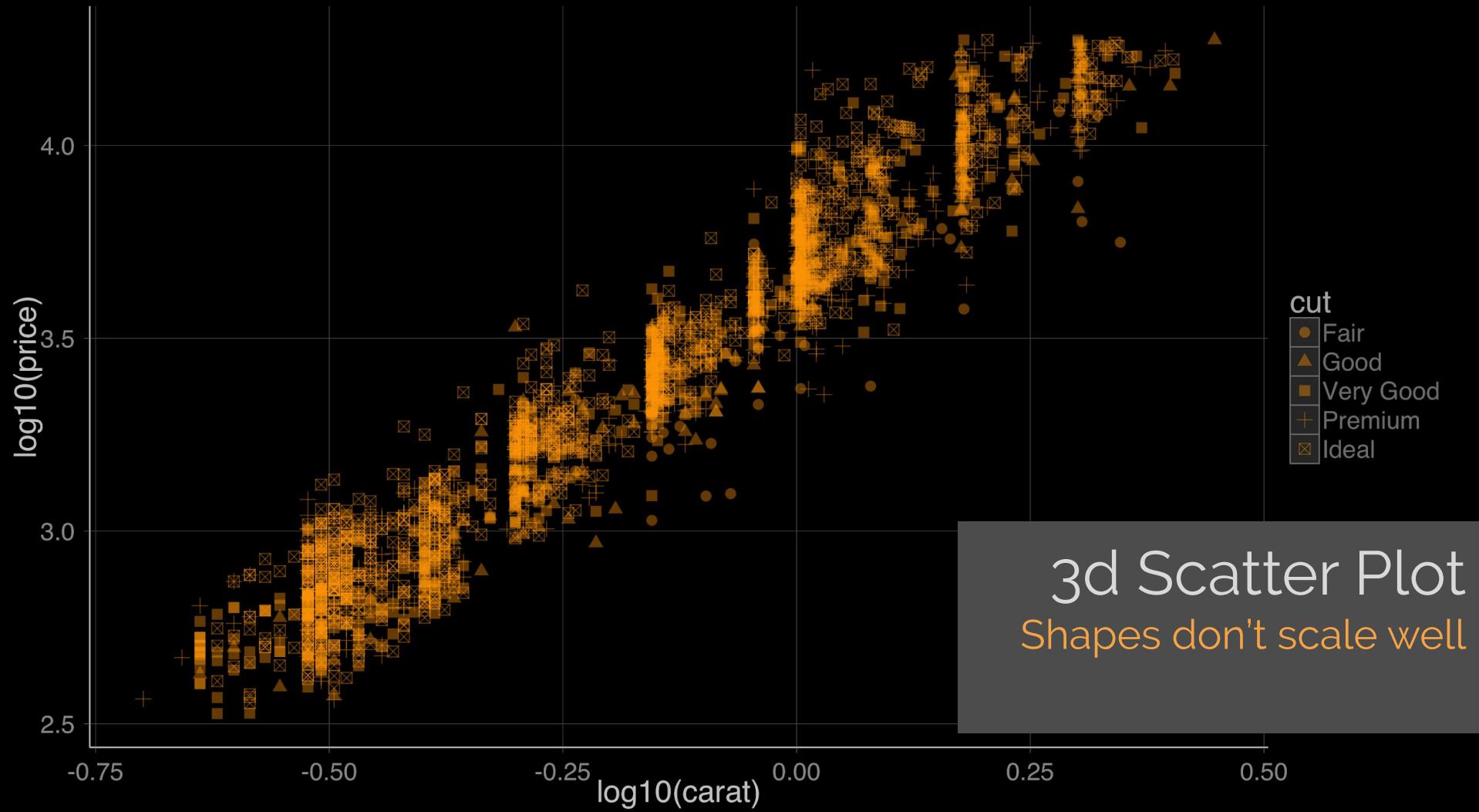
Size

Color

Shape

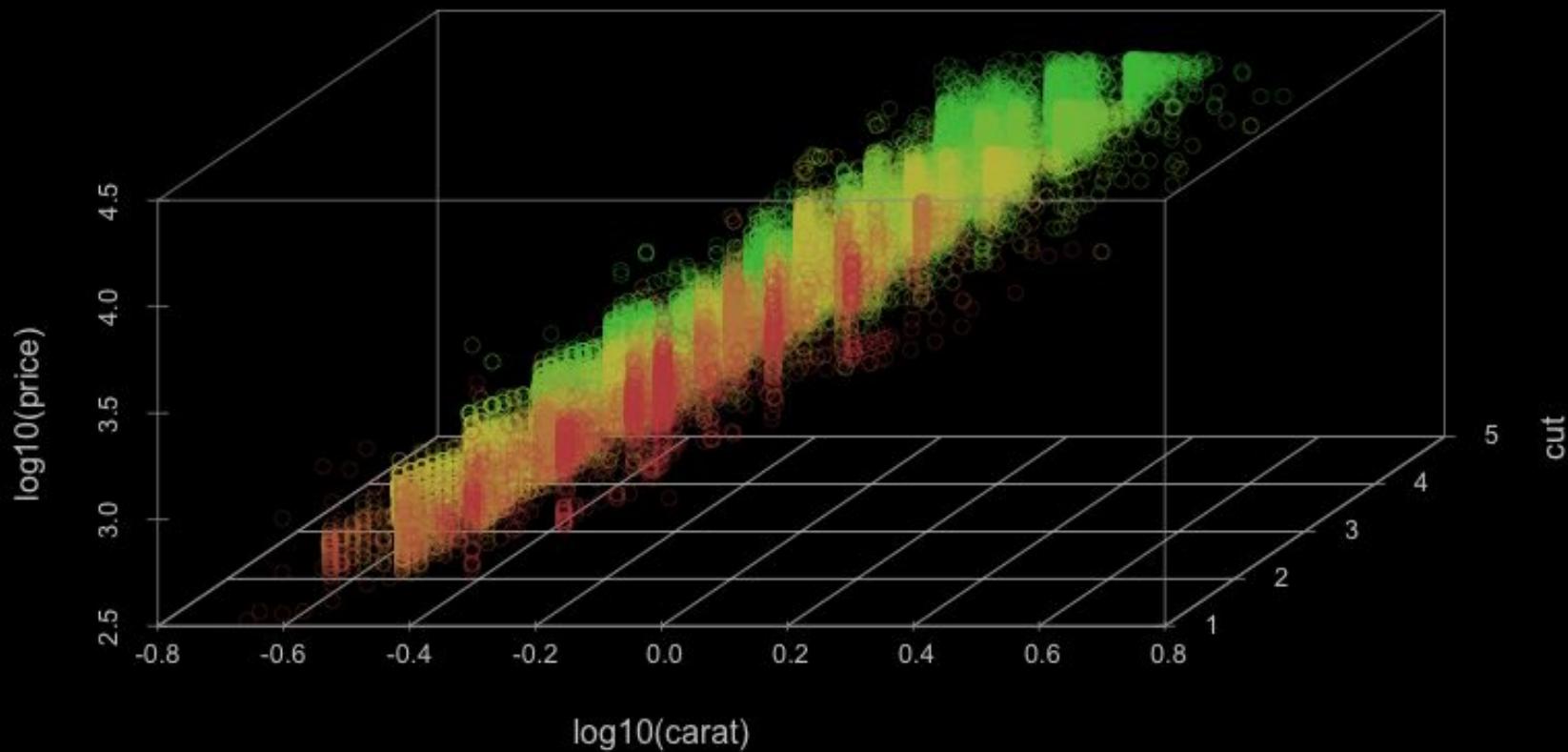






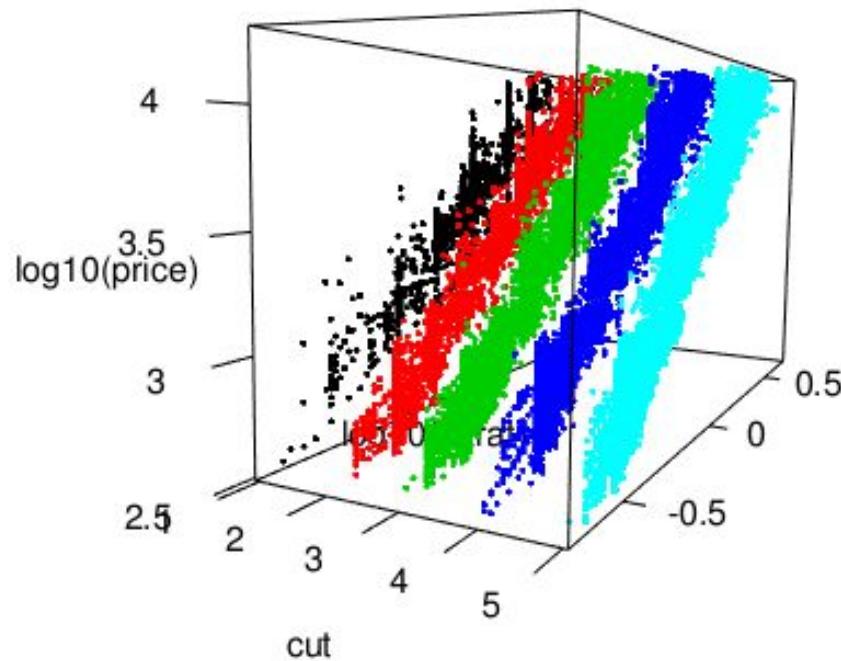
# 3d Scatter Plot

depth persp not good

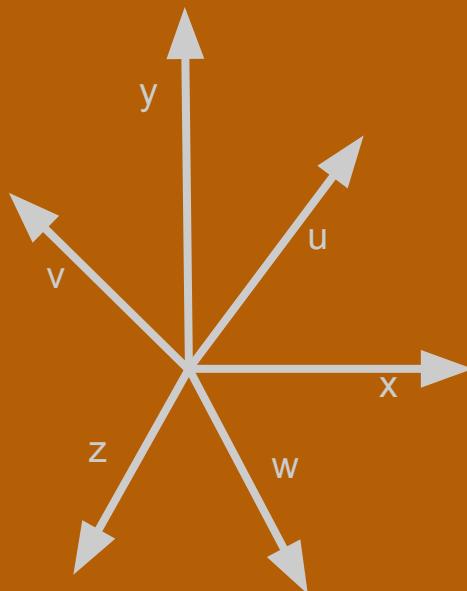


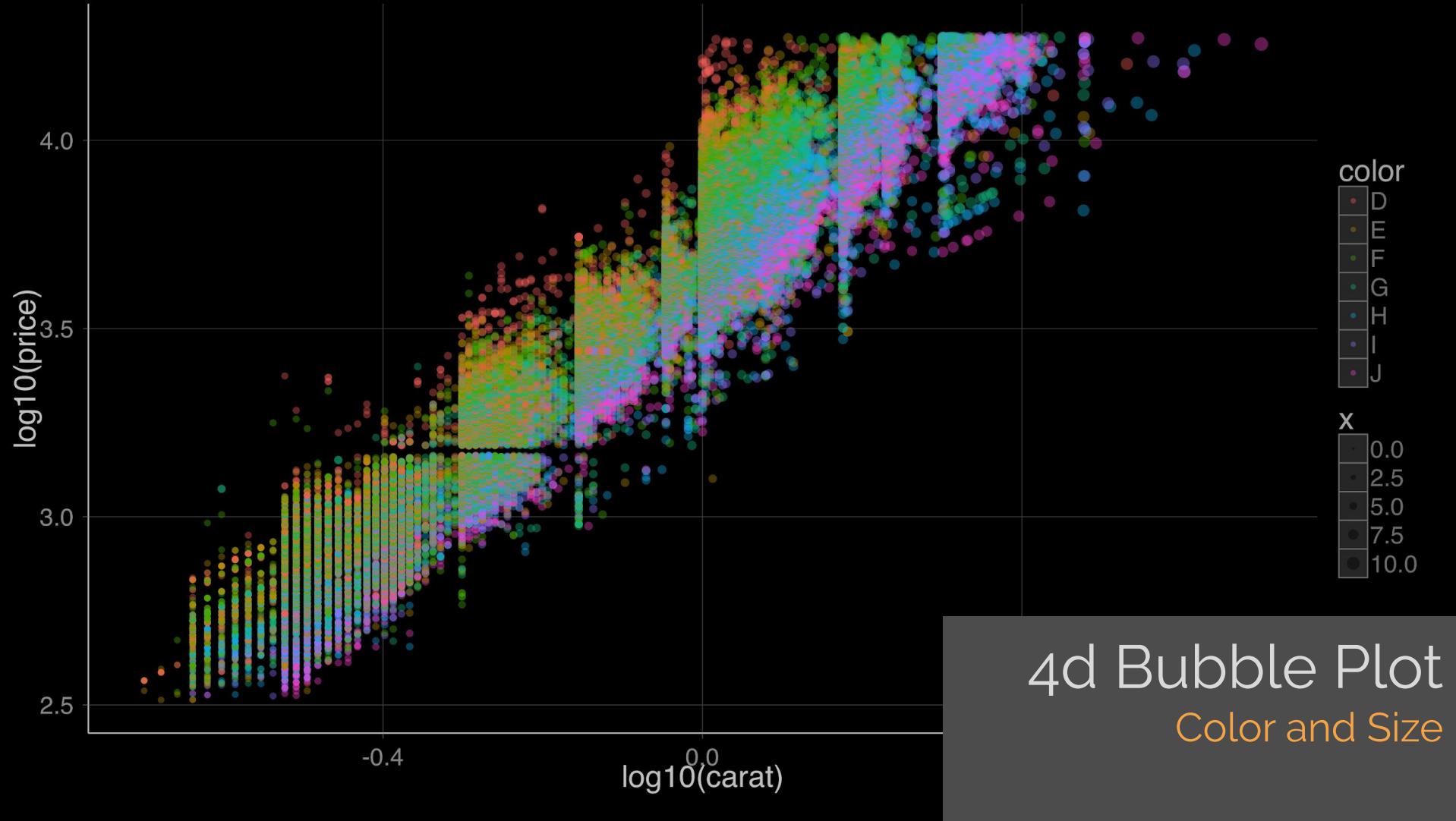
# 3d Scatter Plot

## Interaction: Rotation



4d to 6d



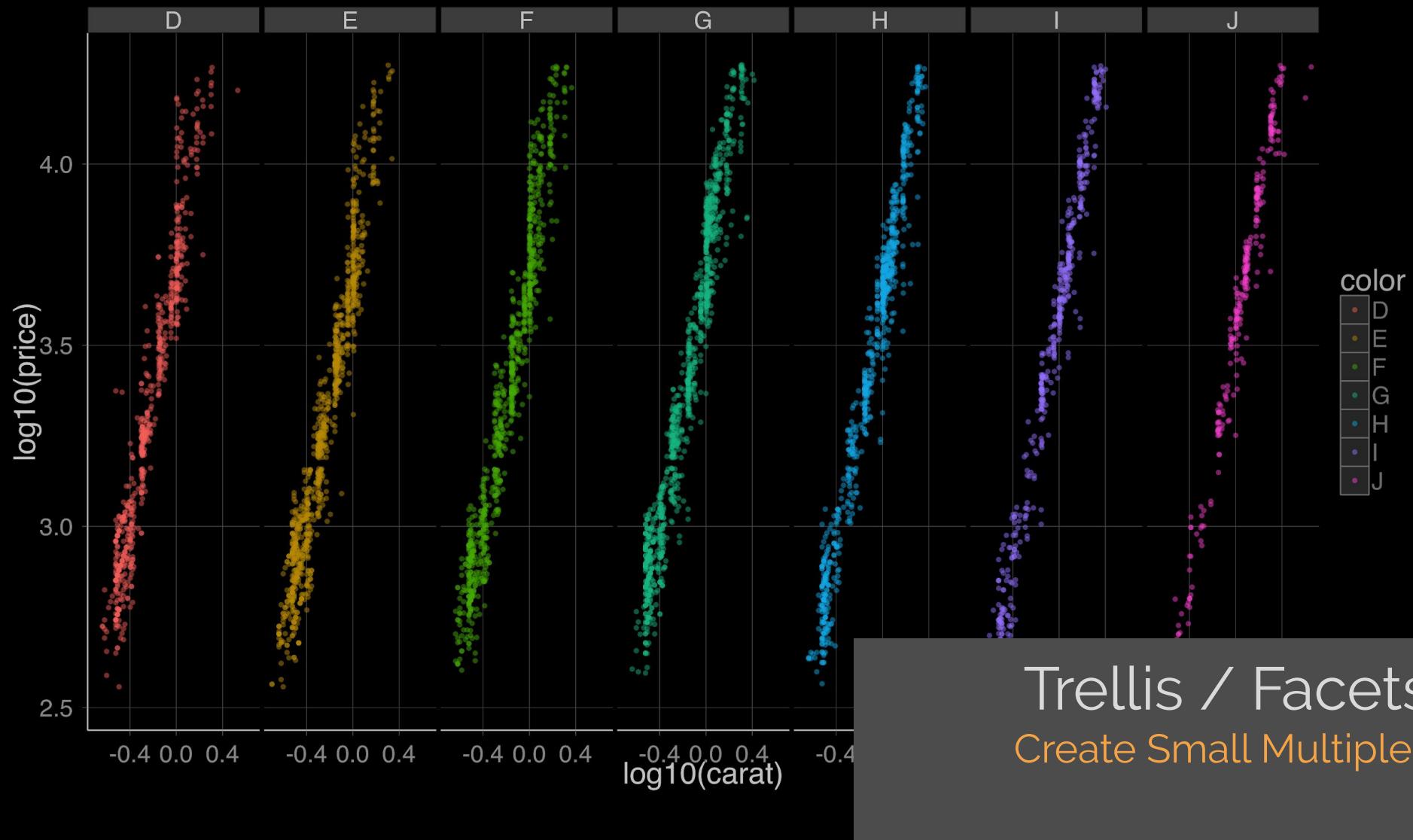


BBC FOUR



5d Bubble Plot  
Color, Size and Time

The Joy of Stat - Hans Rosling



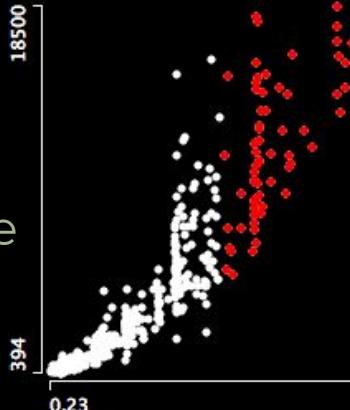


Trellis / Facet Grid  
Create Small Multiples

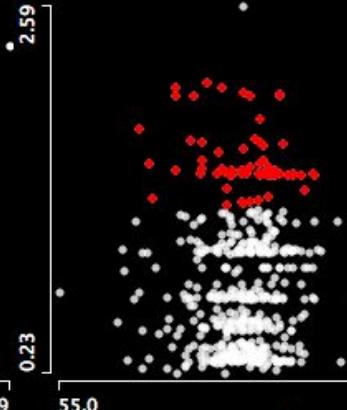
# SPLOM

## Scatterplot Matrix

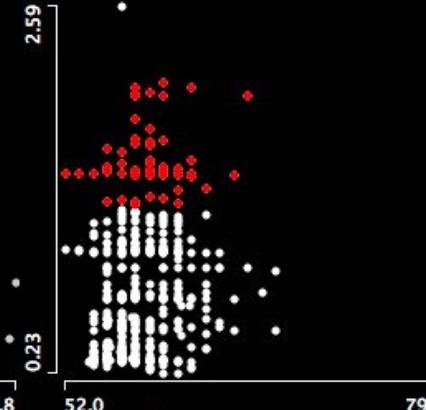
Price



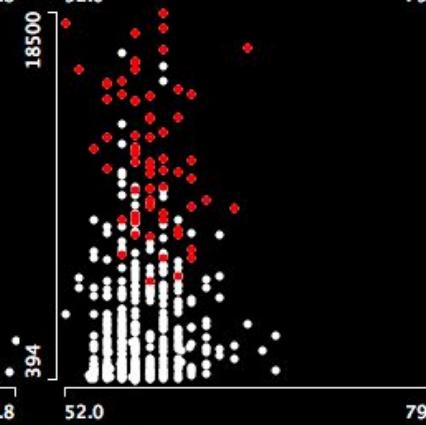
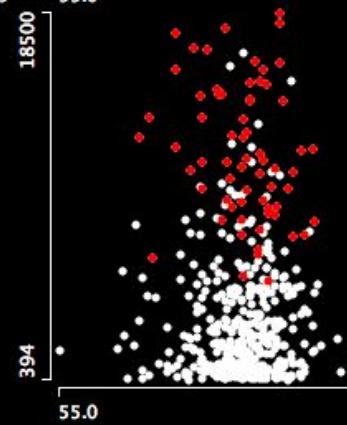
Carat



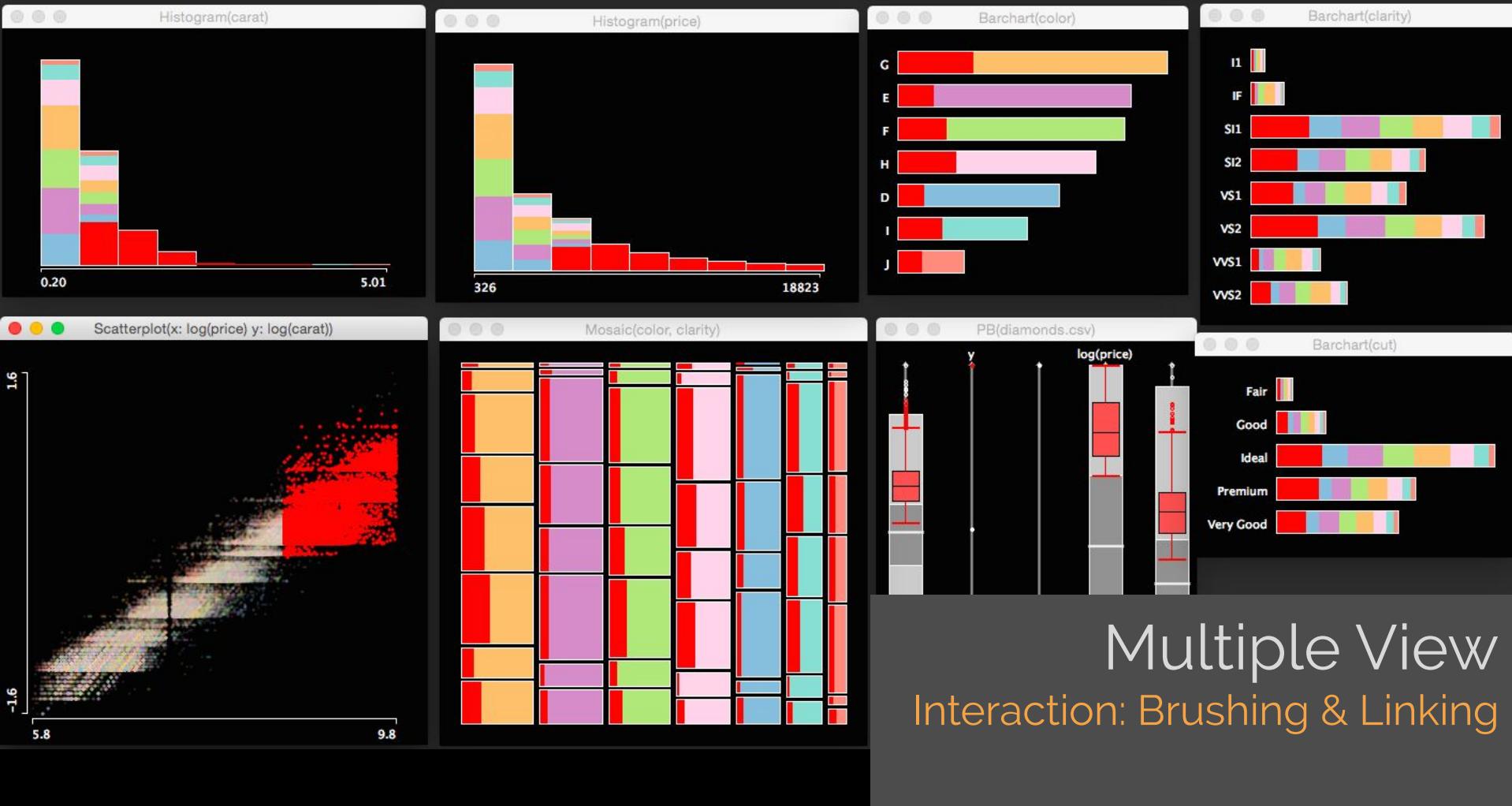
Table



Depth

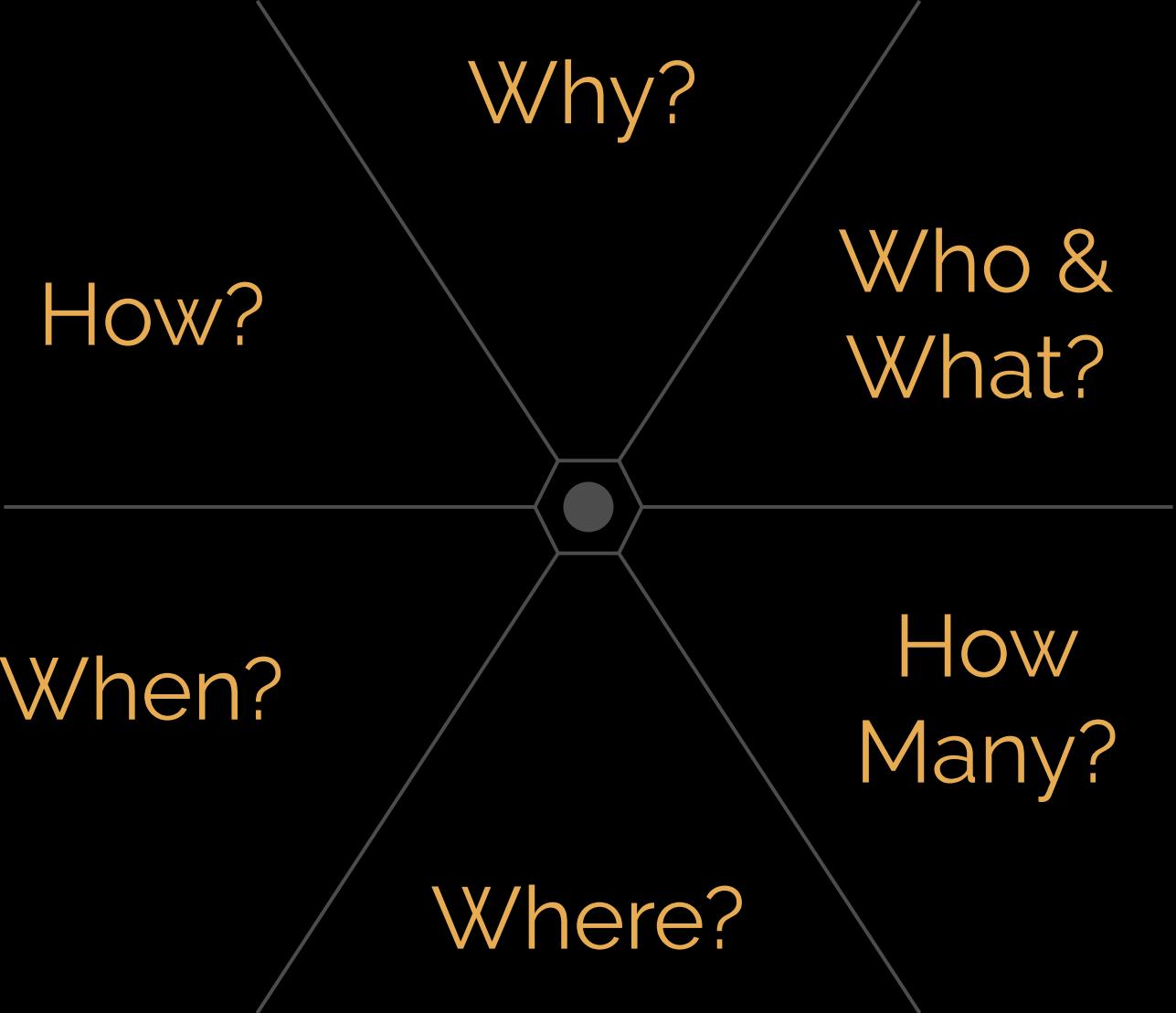






# Choosing Graphs

## “Visualization Zoo”



Why?

How?

Who &  
What?

When?

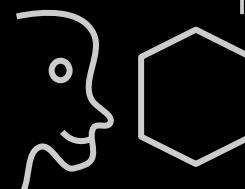
How  
Many?

Where?

# Portrait

Distribution

Representation



Why?

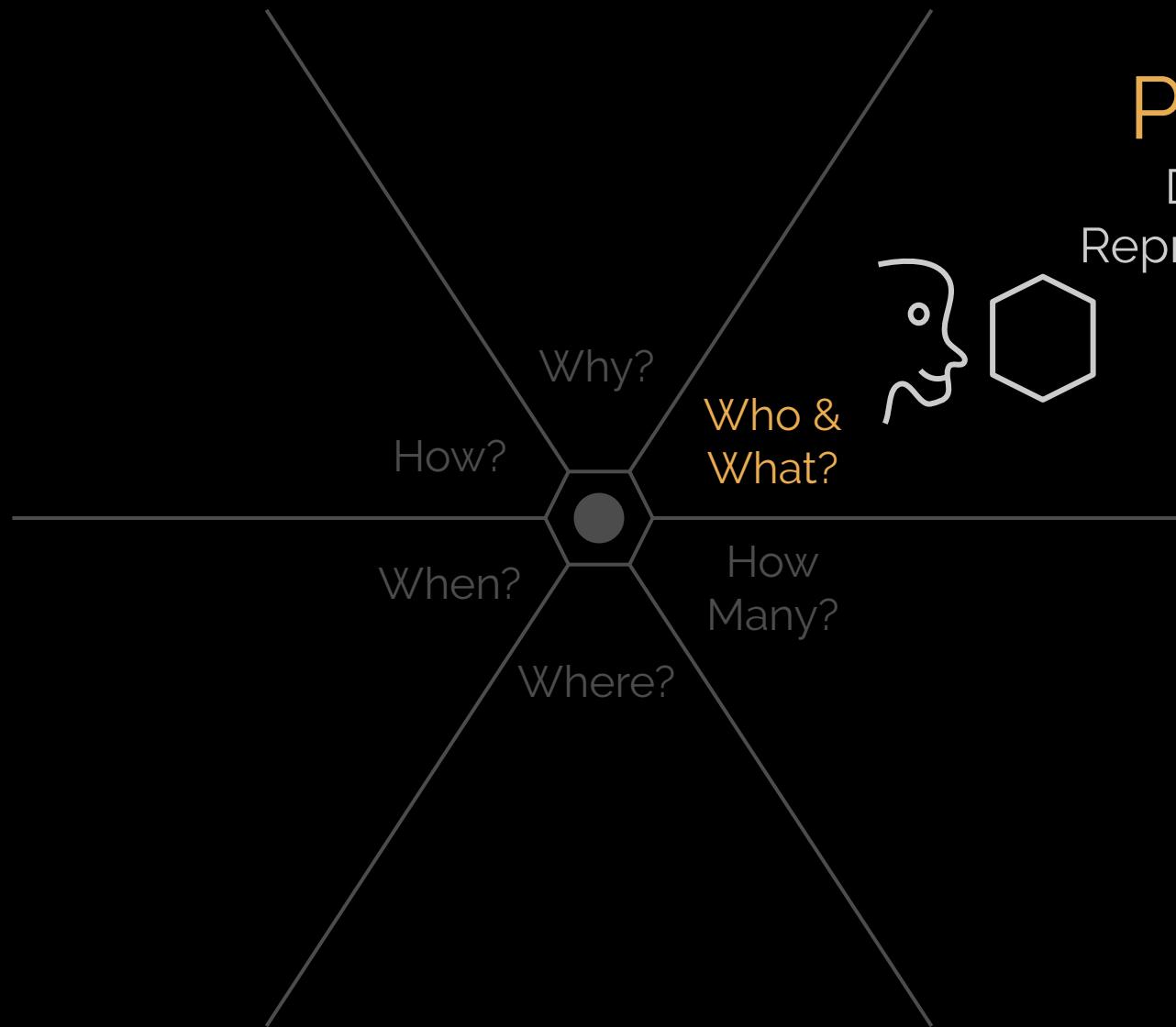
How?

Who &  
What?

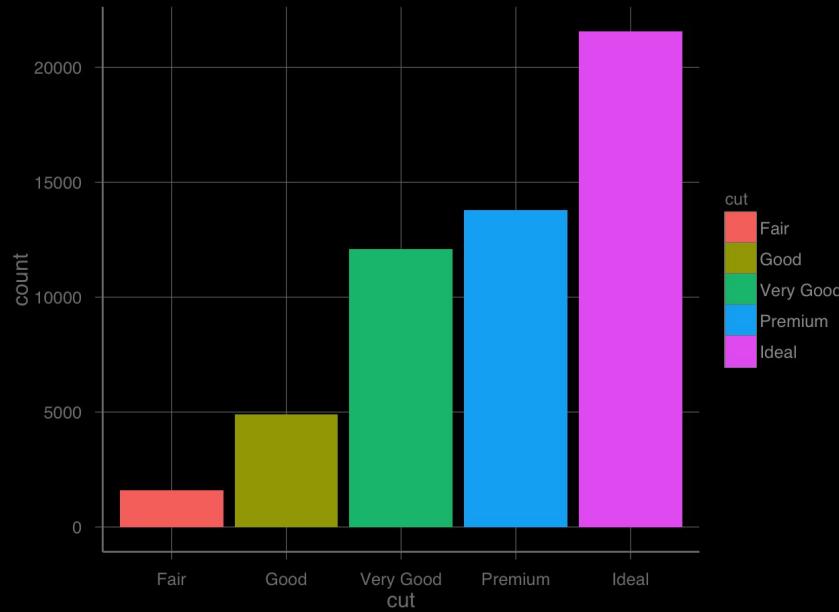
When?

How  
Many?

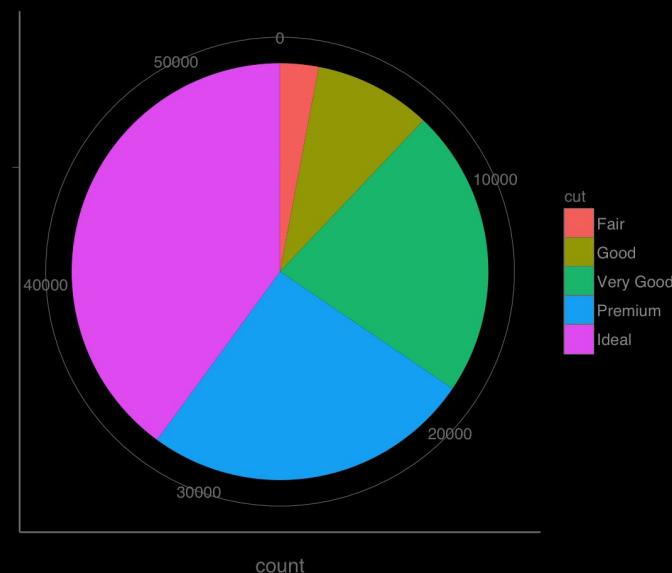
Where?



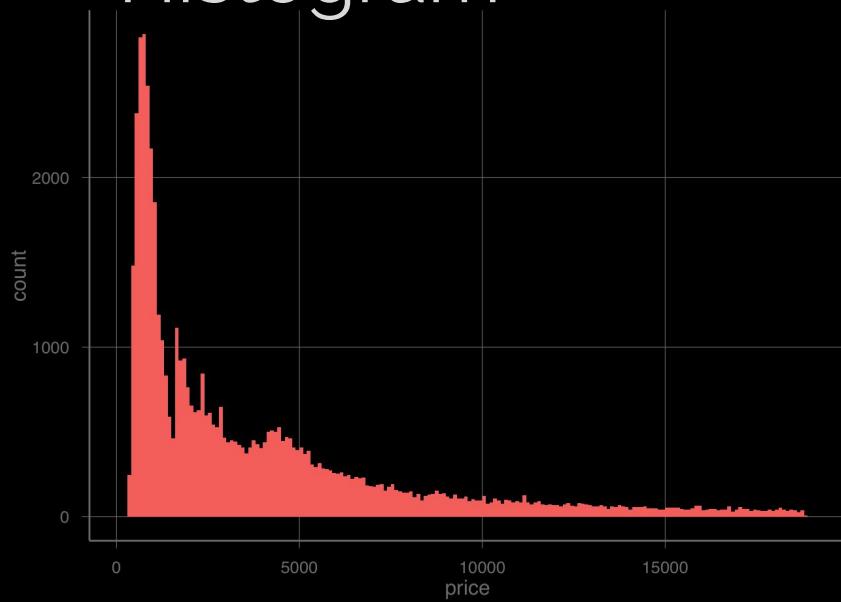
# Column / Bar



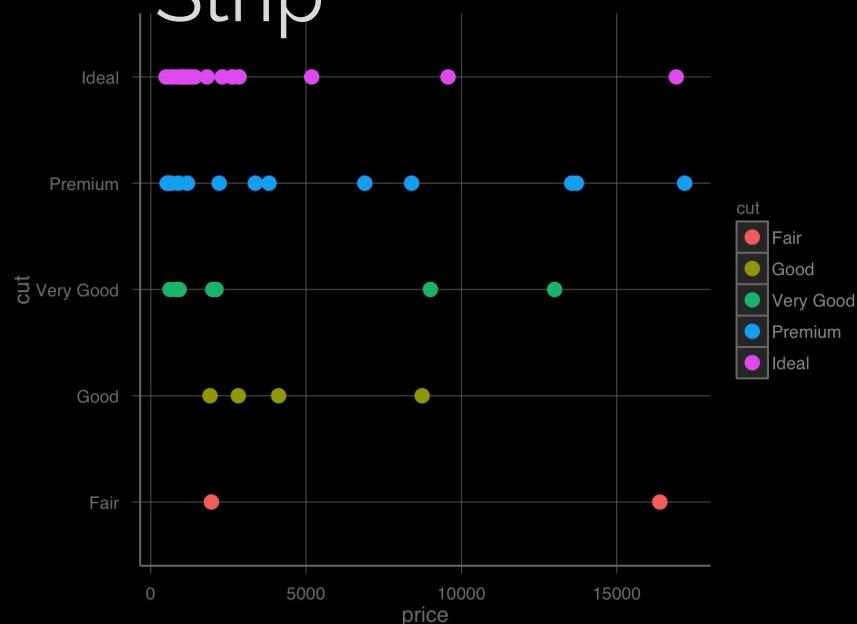
# Pie



# Histogram



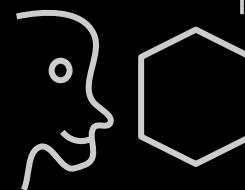
# Strip



Portrait

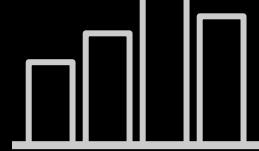
Distribution

Representation



Who &  
What?

How  
Many?



Why?  
How?

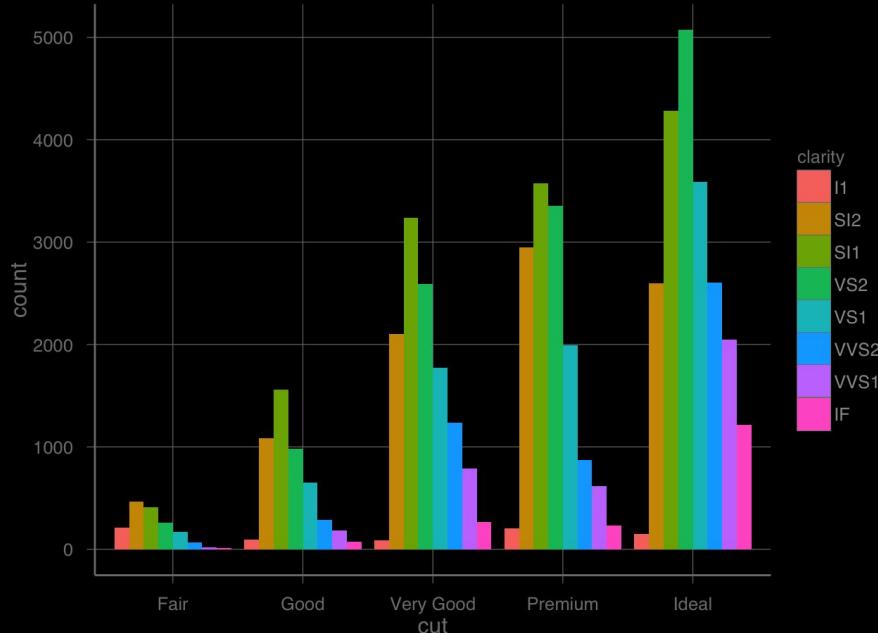
When?

Where?

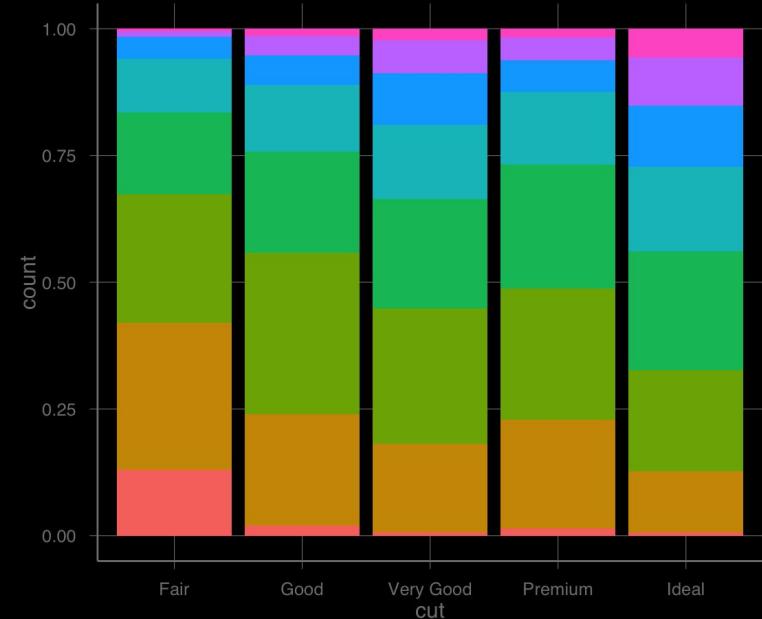
Comparison

Comparative  
Representation

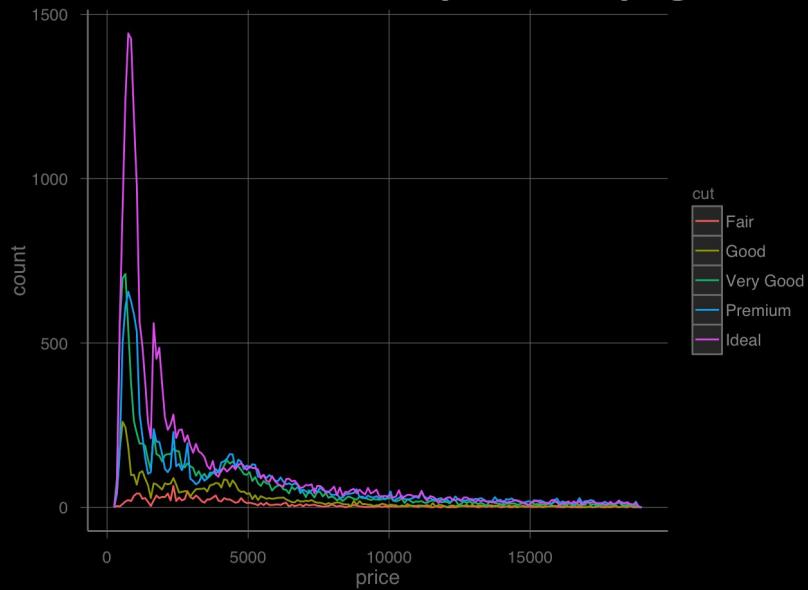
# Column / Bar

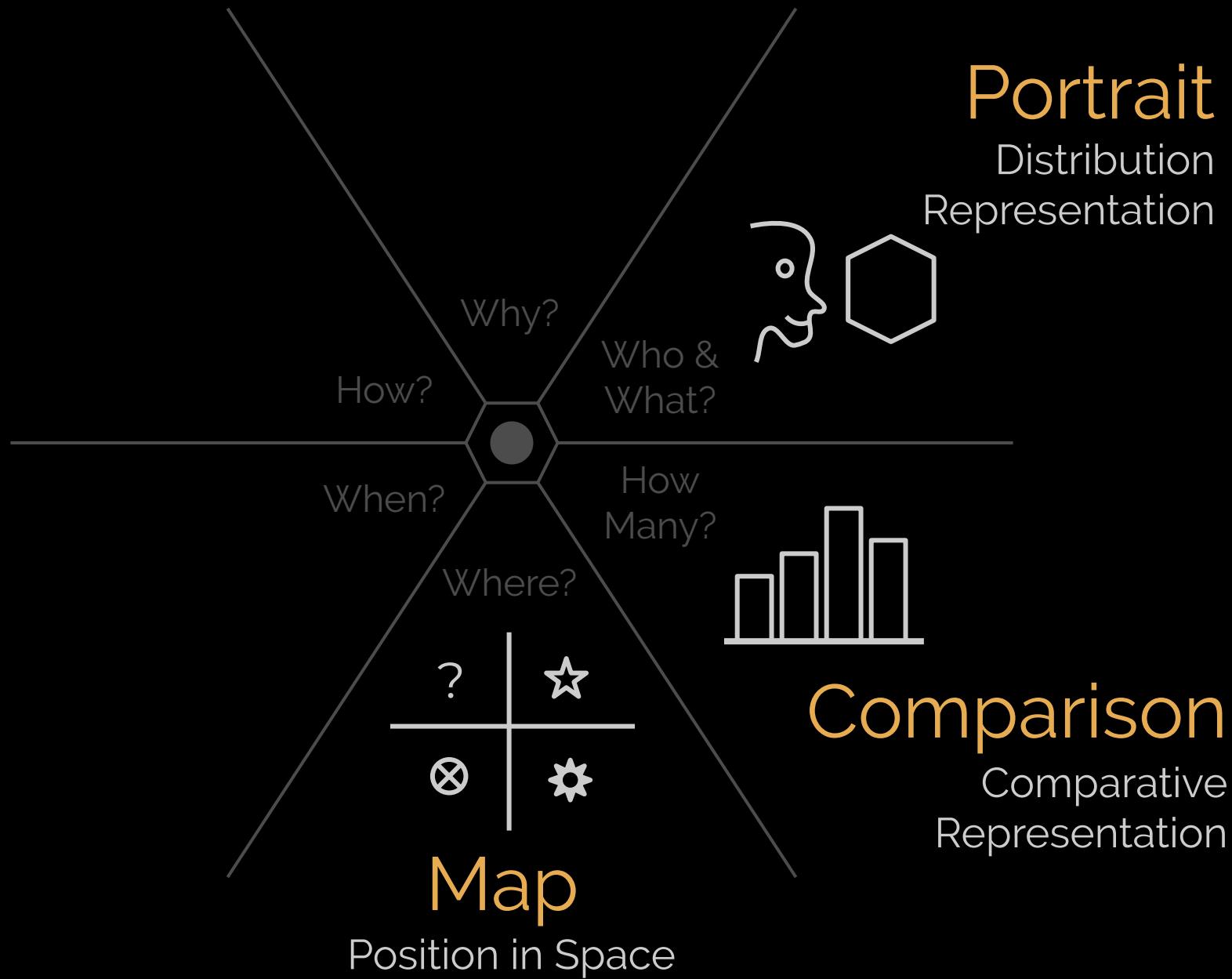


# Stacked



# Frequency Polygon



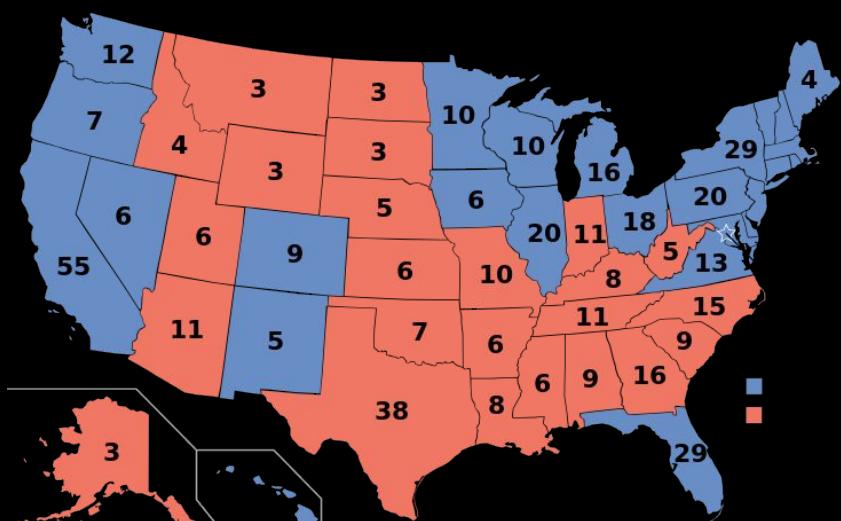


# Understanding Maps



Interactive Demo : [Map Projections Examples](#)

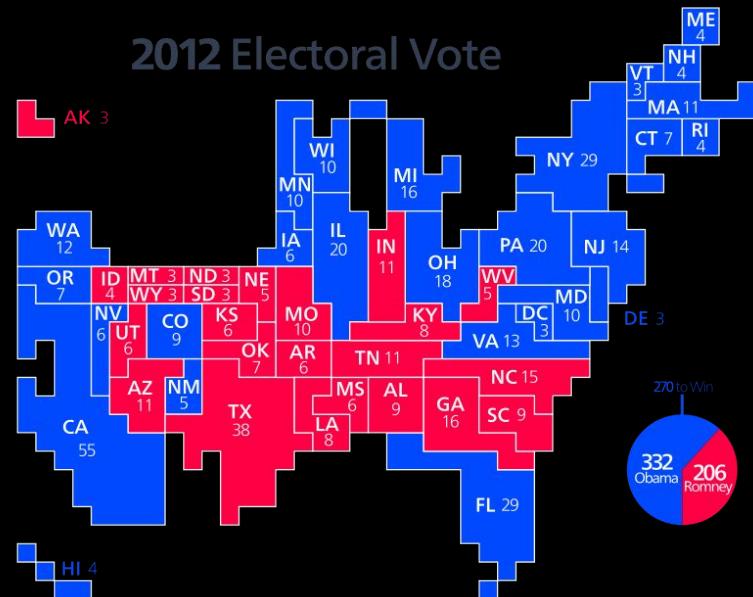
# Chloropleth



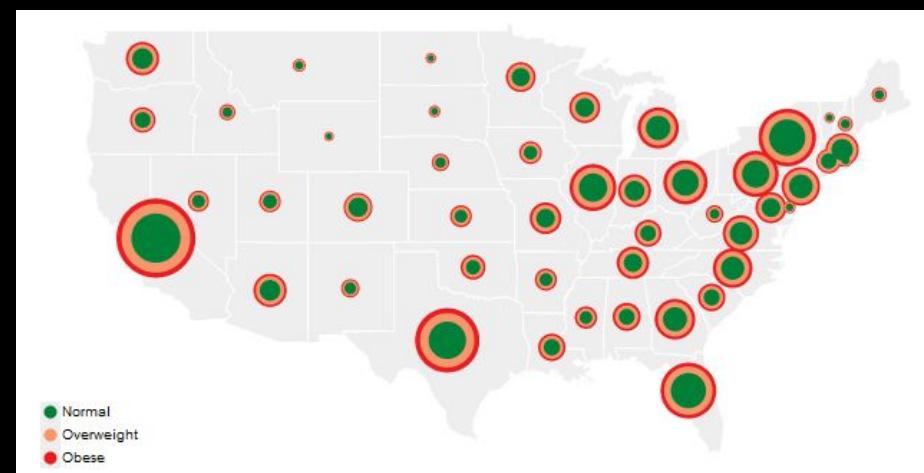
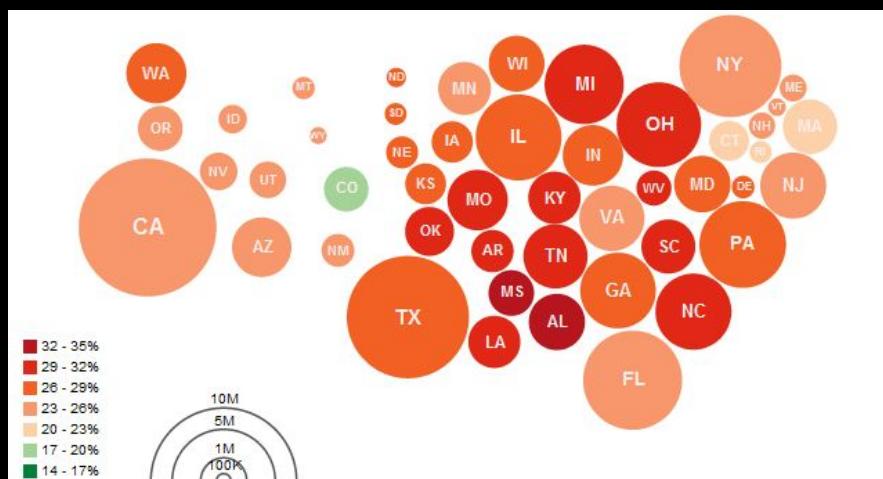
Dorling  
Cartogram

# Cartogram

## 2012 Electoral Vote



Graduated

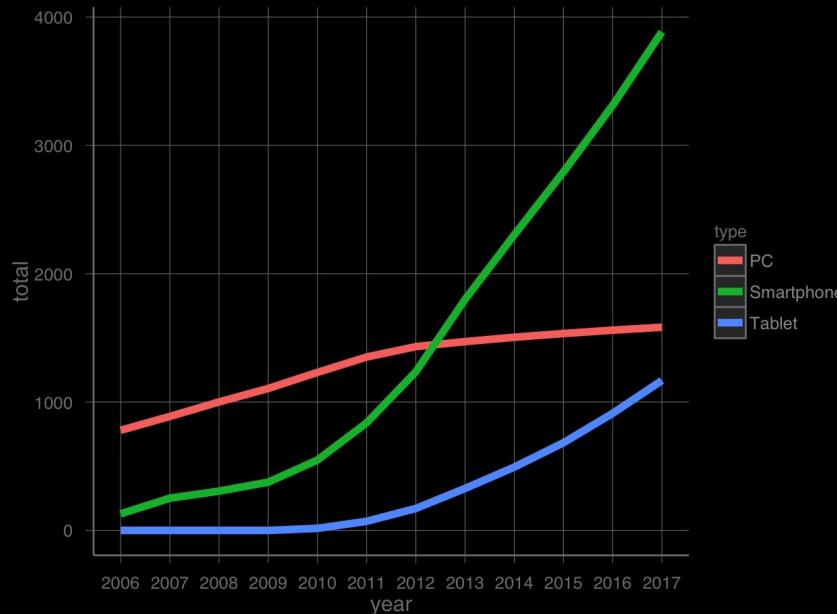


**Timeline**  
Position in Time

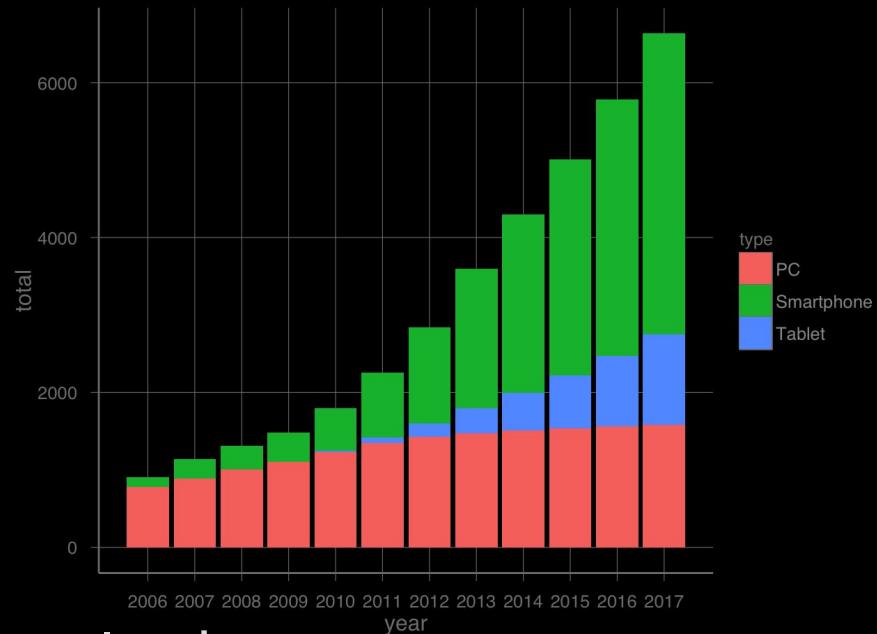


Position in Space

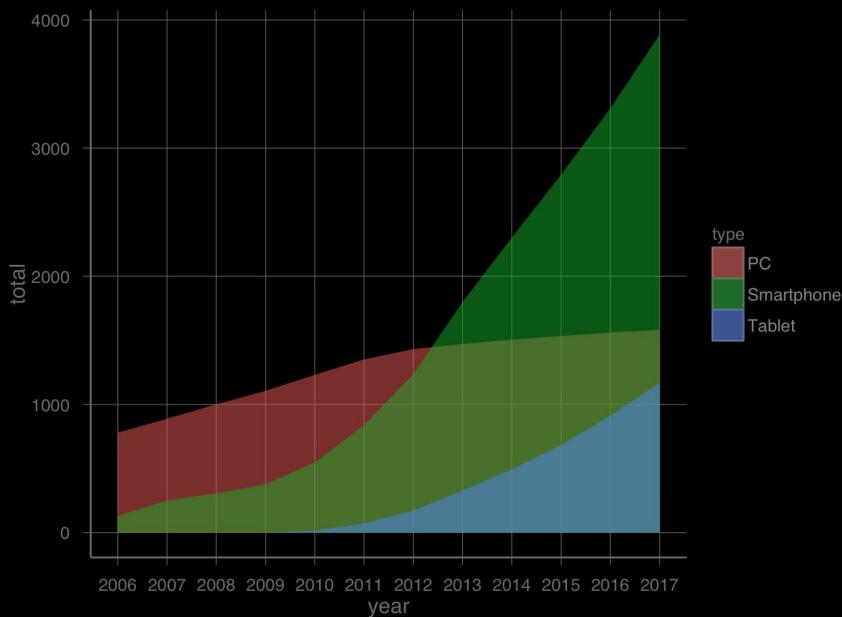
# Line



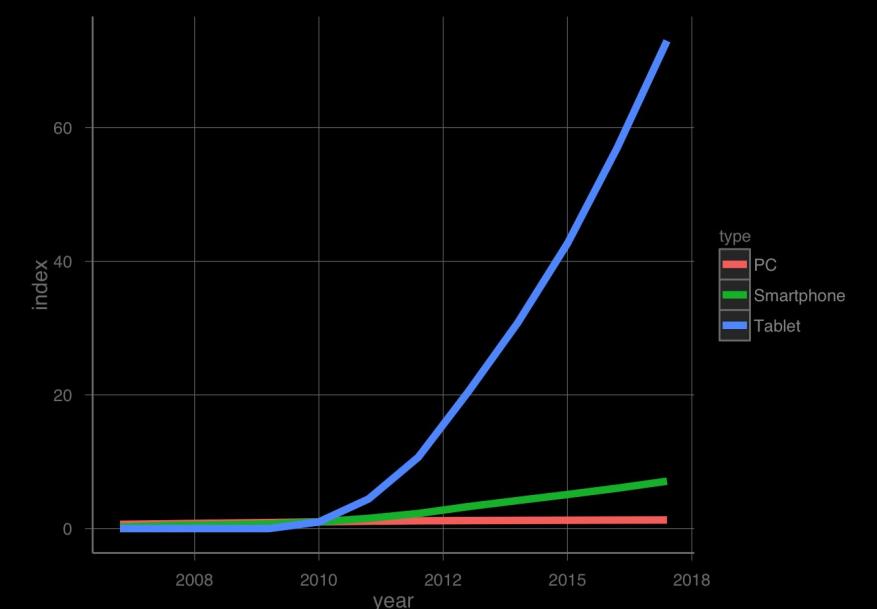
# Stacked



# Horizon

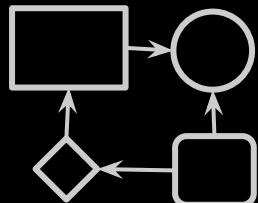


# Index

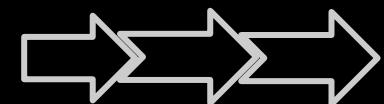


# Flowchart

Relationship,  
Hierarchy



Why?  
How?

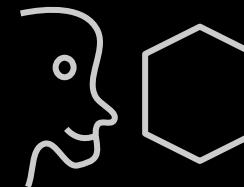


# Timeline

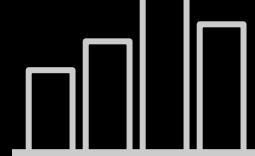
Position in  
Time

# Portrait

Distribution  
Representation



Who &  
What?

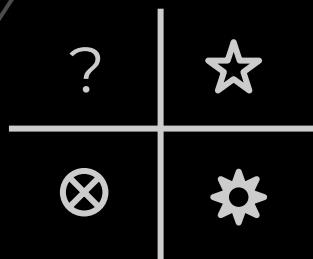


# Comparison

Comparative  
Representation

# Map

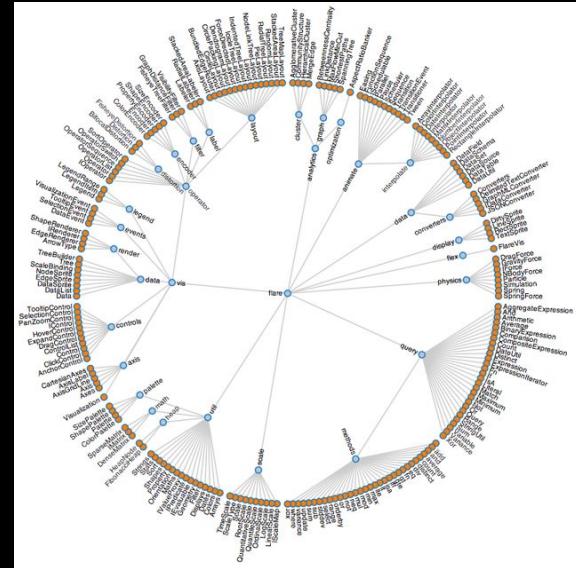
Position in Space



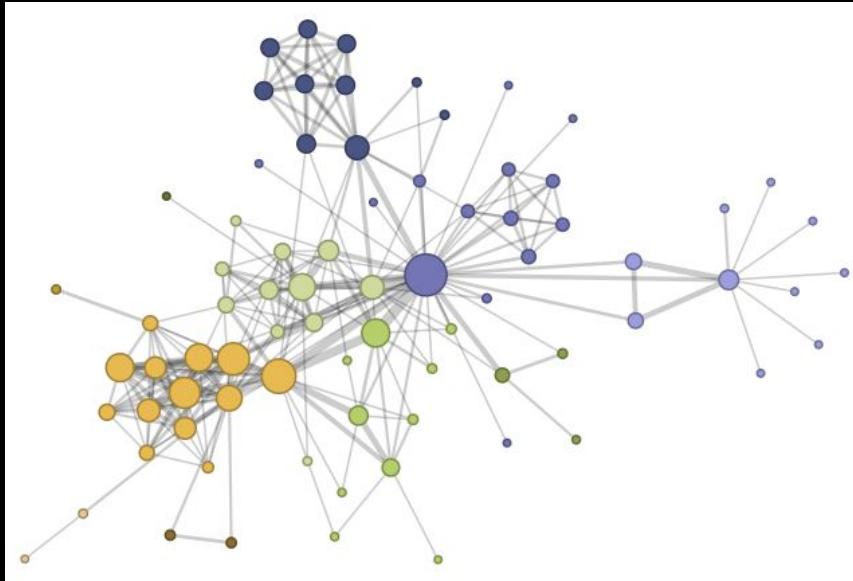
# Enclosure



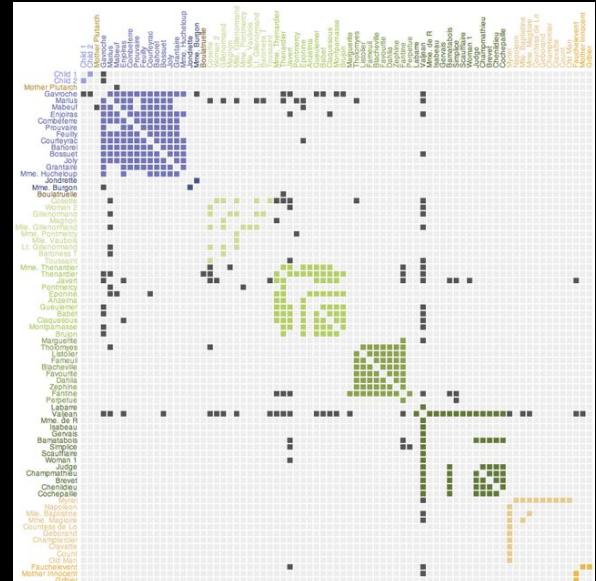
# Tree Radial



# Force Directed



## Matrix

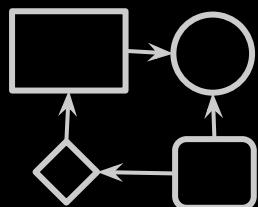


# Multi Variable

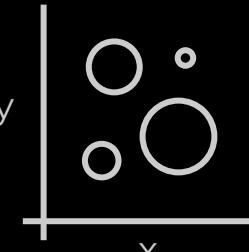
Deduction & Prediction

## Flowchart

Relationship,  
Hierarchy



How?

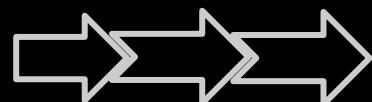


Why?

When?

## Timeline

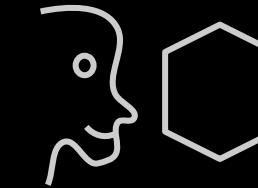
Position in  
Time



Where?

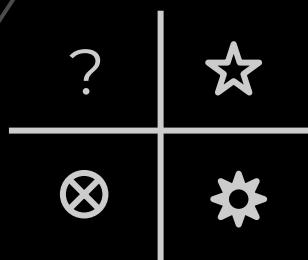
How  
Many?

Who &  
What?



## Portrait

Distribution  
Representation



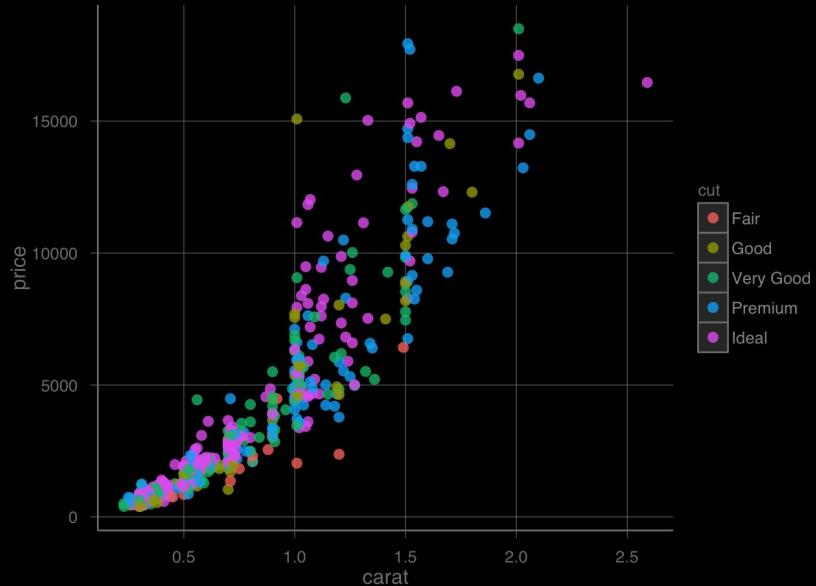
## Comparison

Comparative  
Representation

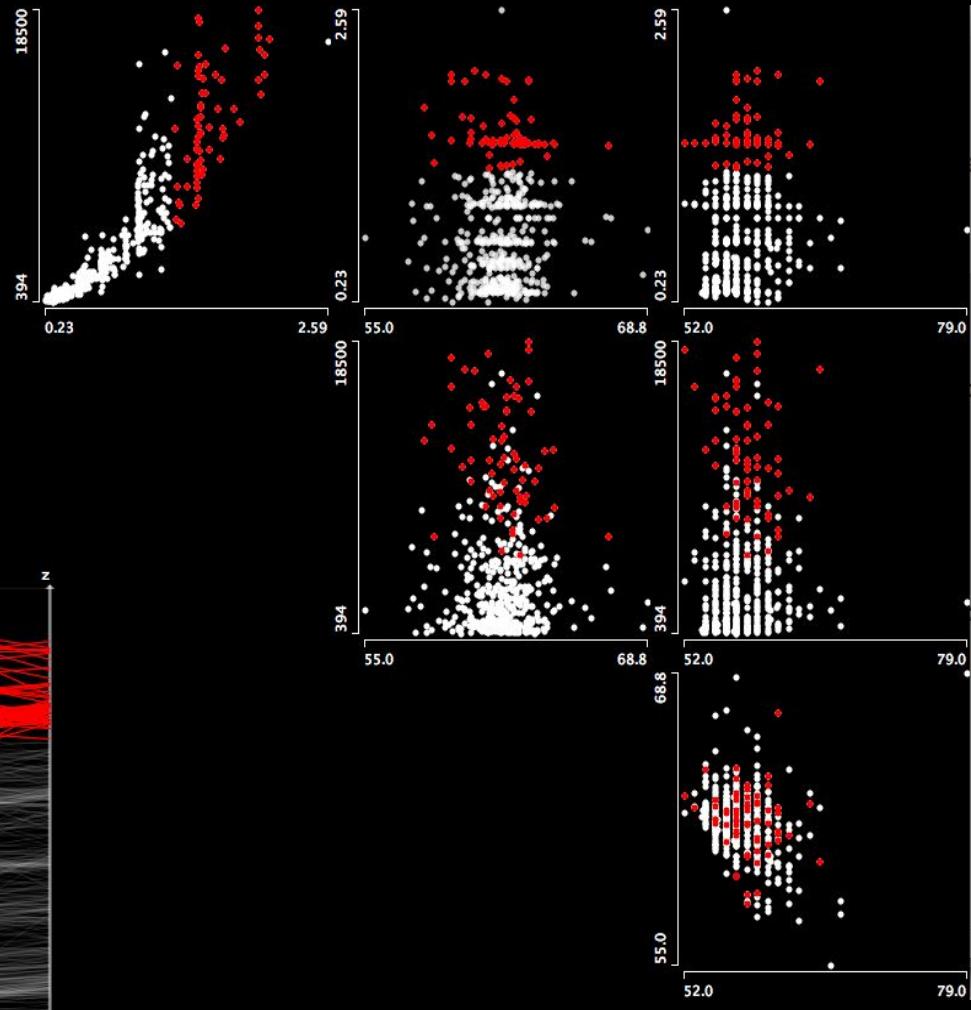
## Map

Position in Space

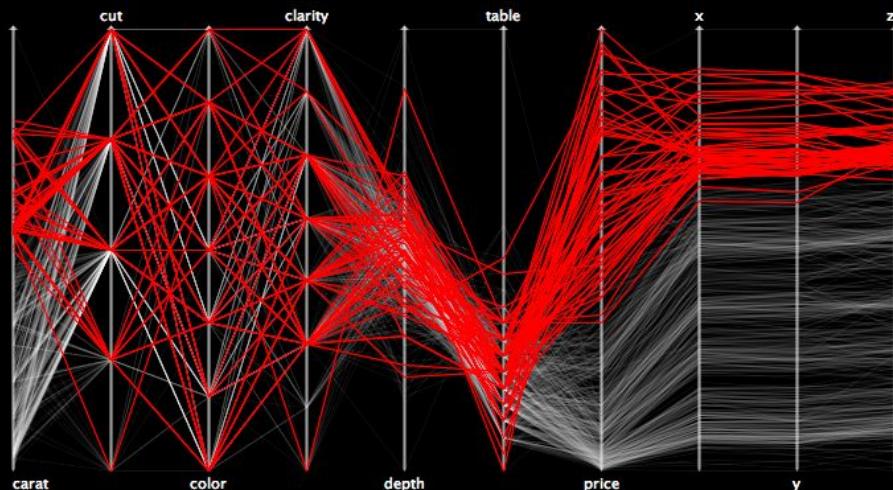
# Scatter & Bubbles



# Scatter Plot Matrix



# Parallel Coordinates

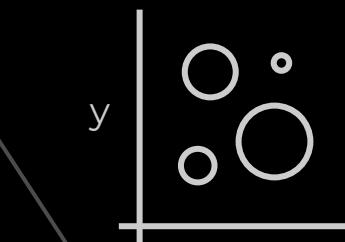
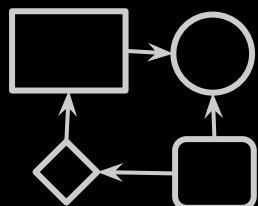


# Multi Variable

Deduction & Prediction

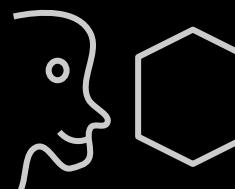
## Flowchart

Relationship,  
Hierarchy



## Portrait

Distribution  
Representation



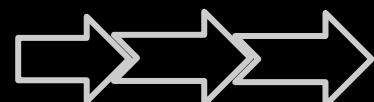
Why?  
How?

Who &  
What?

When?

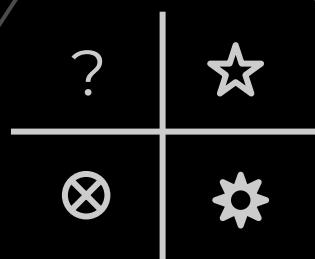
How  
Many?

Where?



## Timeline

Position in  
Time



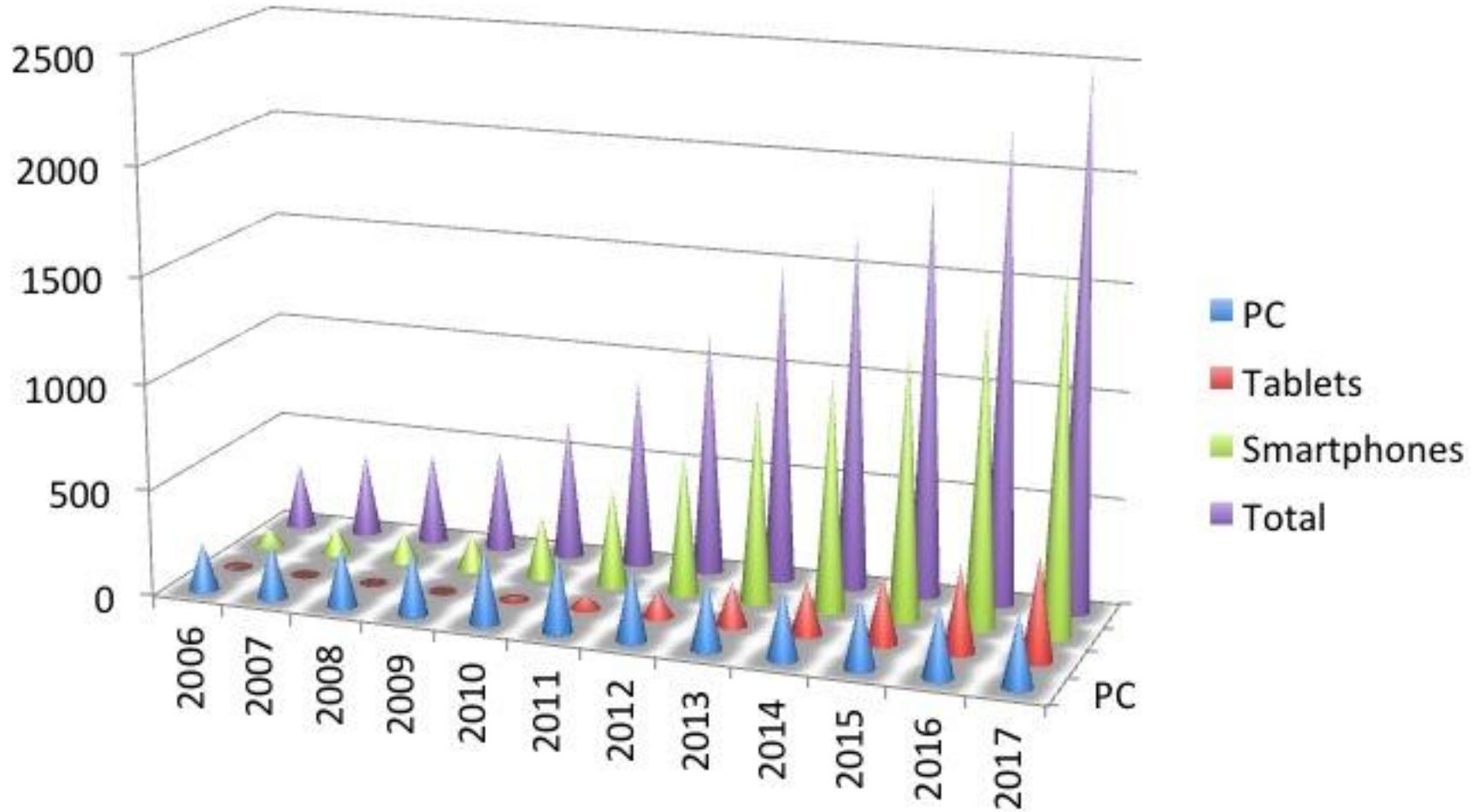
## Comparison

Comparative  
Representation

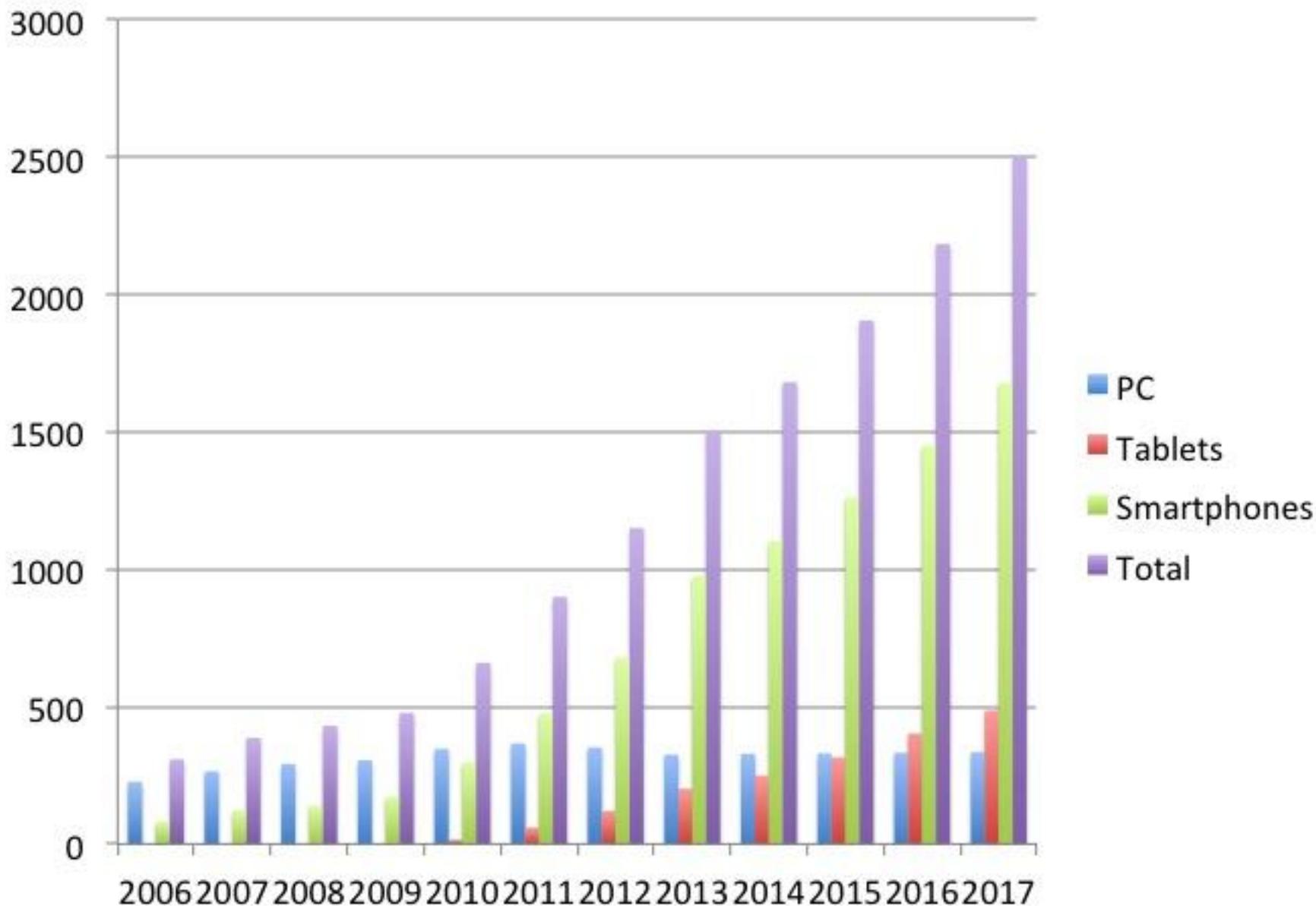
## Map

Position in Space

My chart selection process  
is a lot iterative!

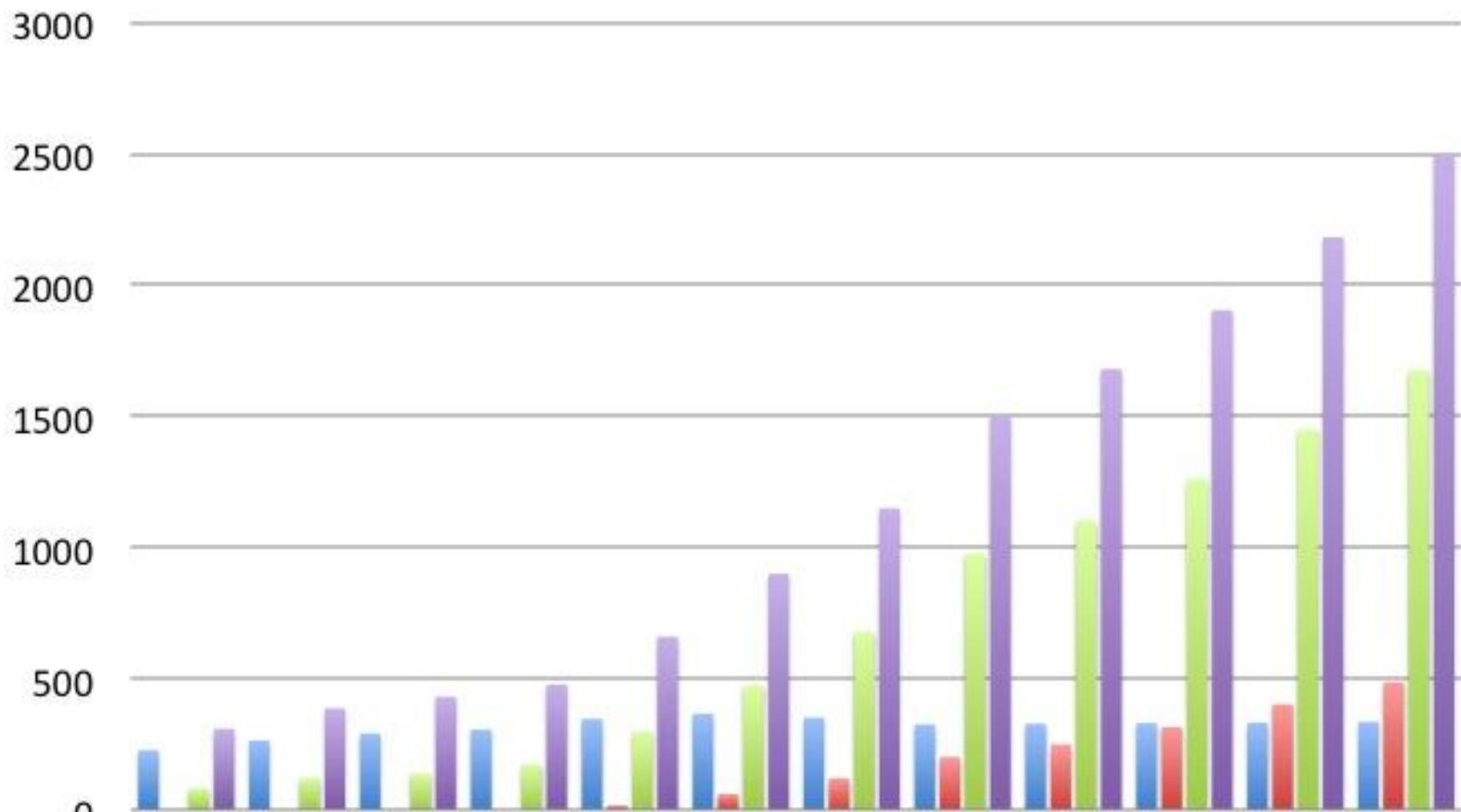


# Standard are Bad?



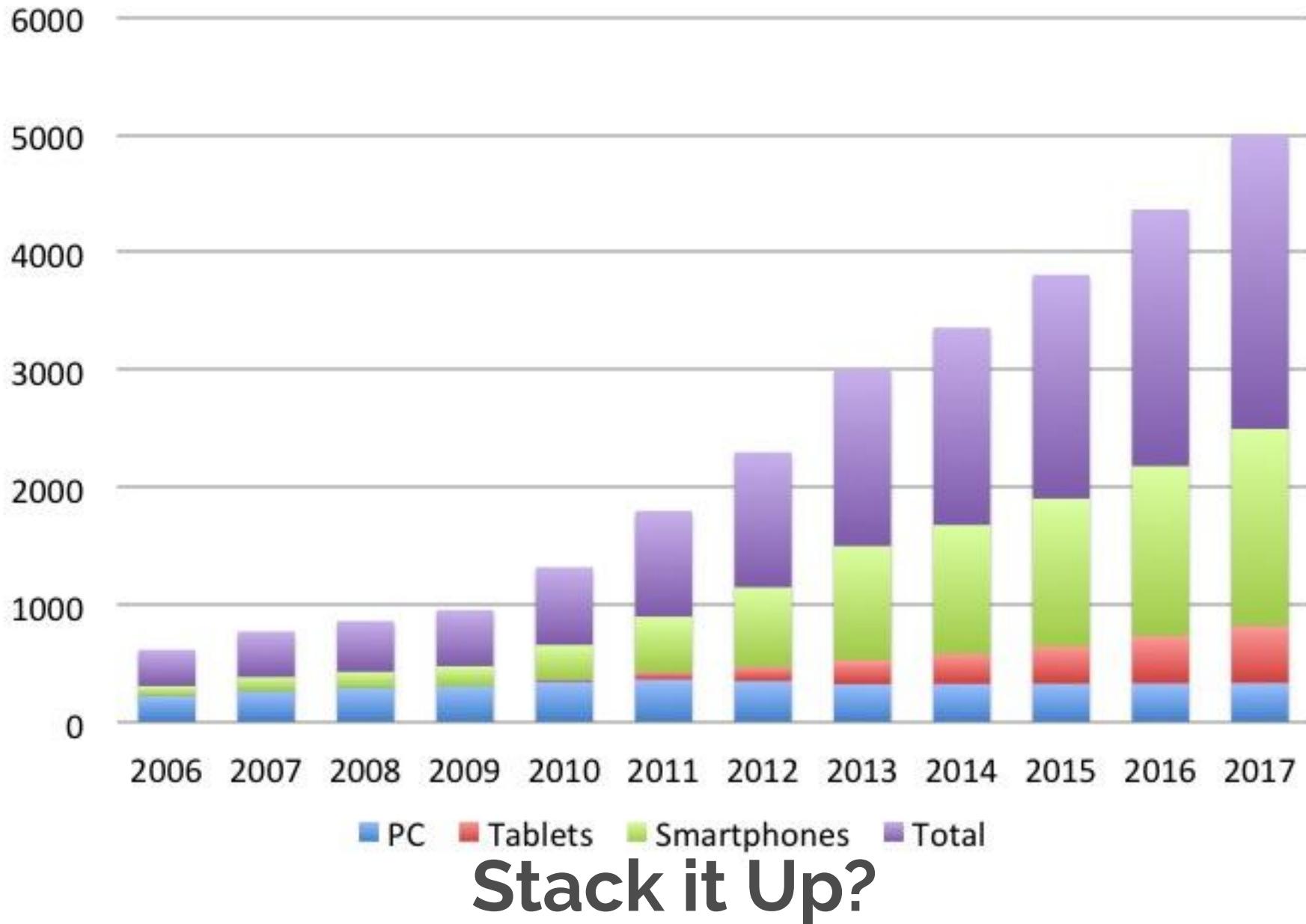
Make it a Bar?

## Device Shipment by Type

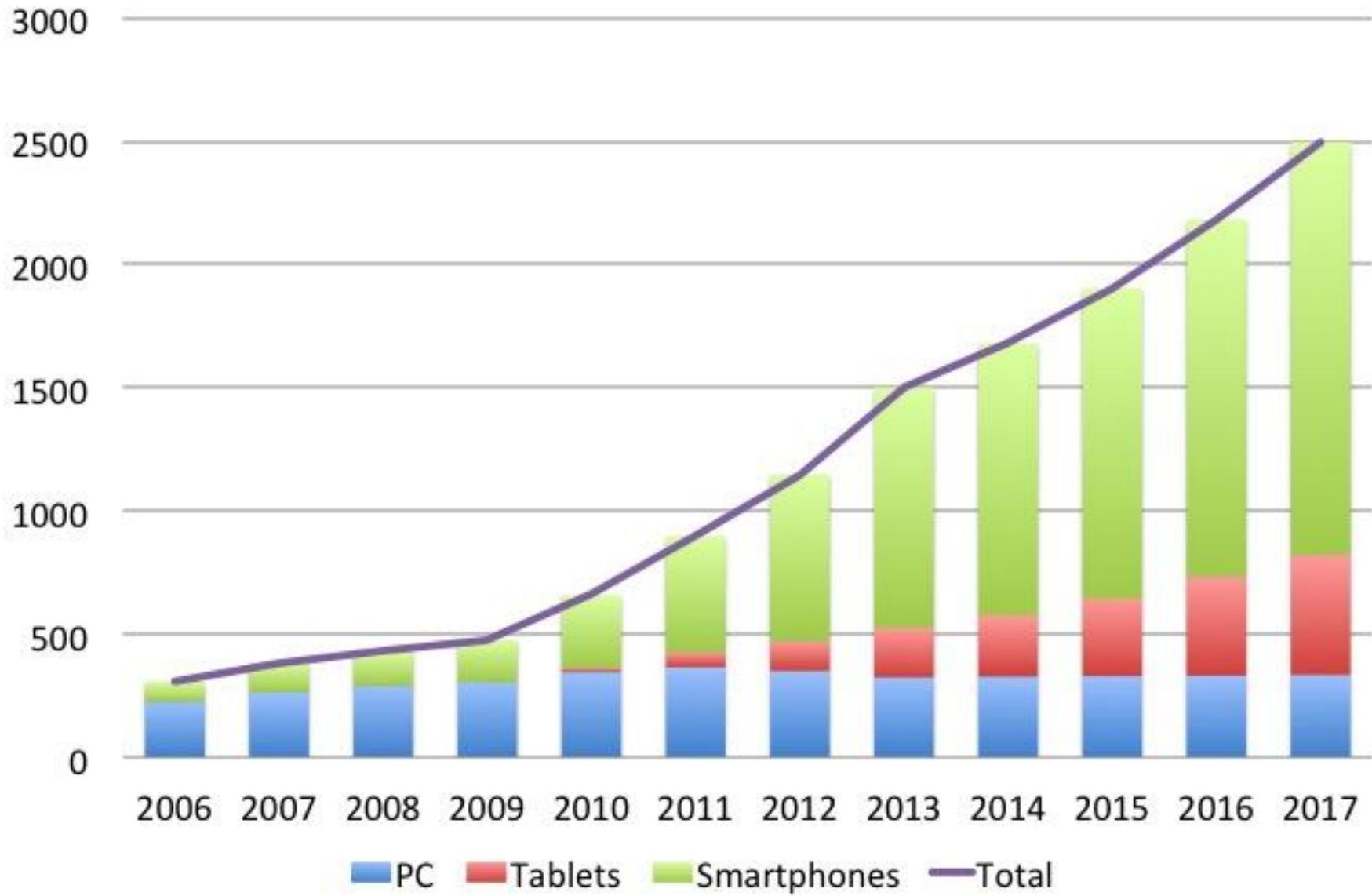


■ PC ■ Tablets ■ Smartphones ■ Total  
**Give it Space?**

## Device Shipment by Type

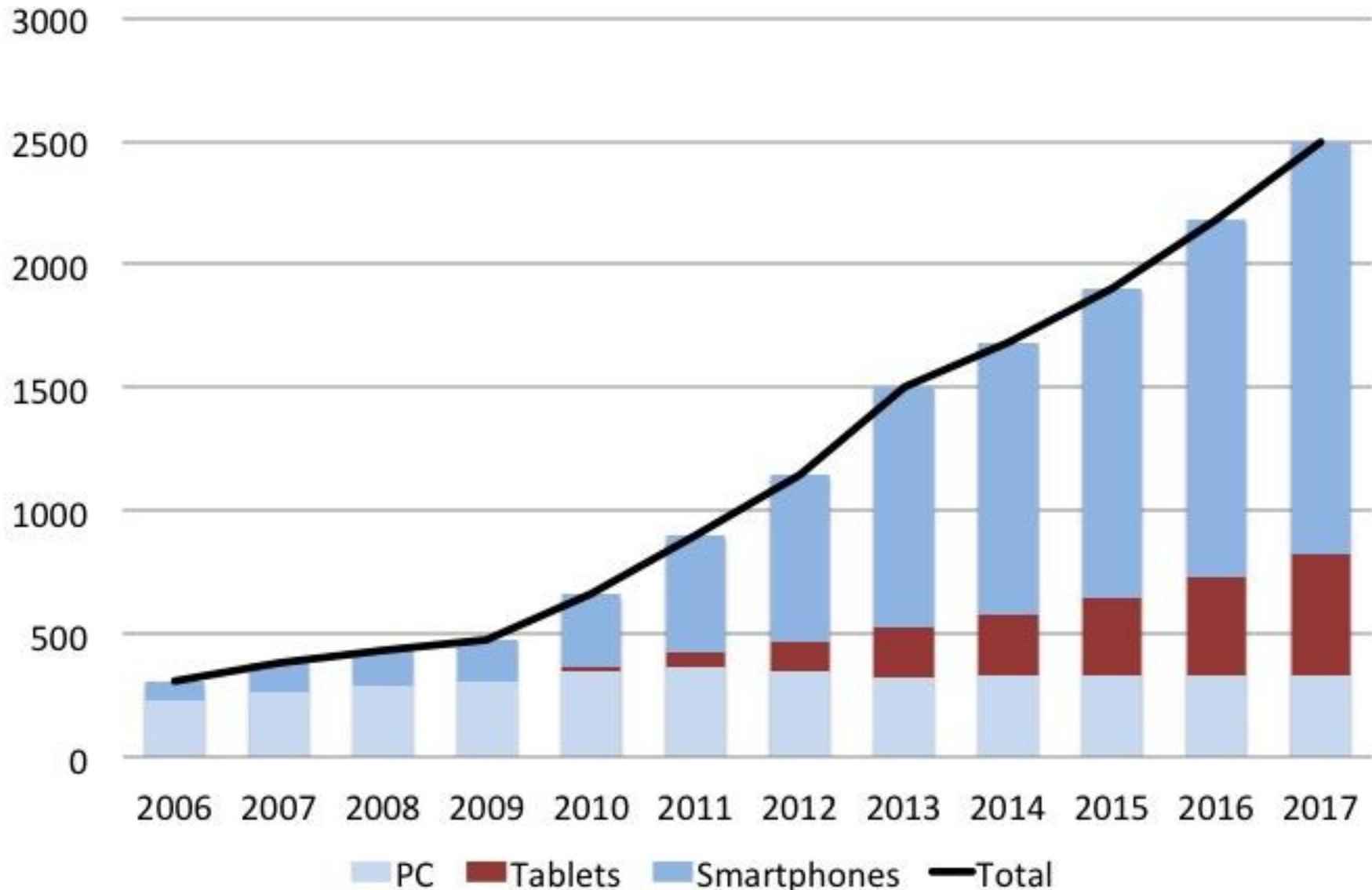


## Device Shipment by Type



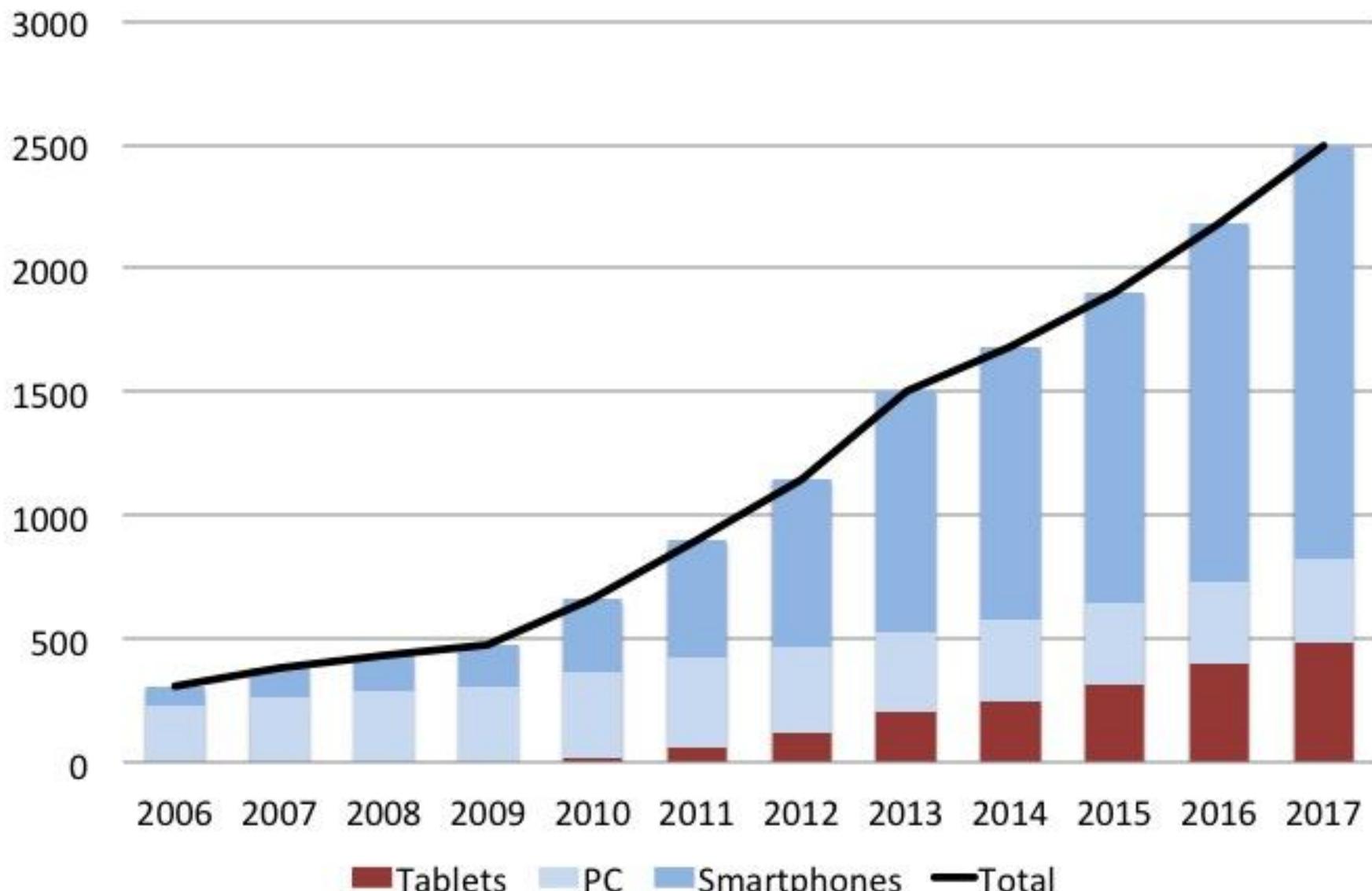
**Remove the Redundant?**

## Device Shipment by Type



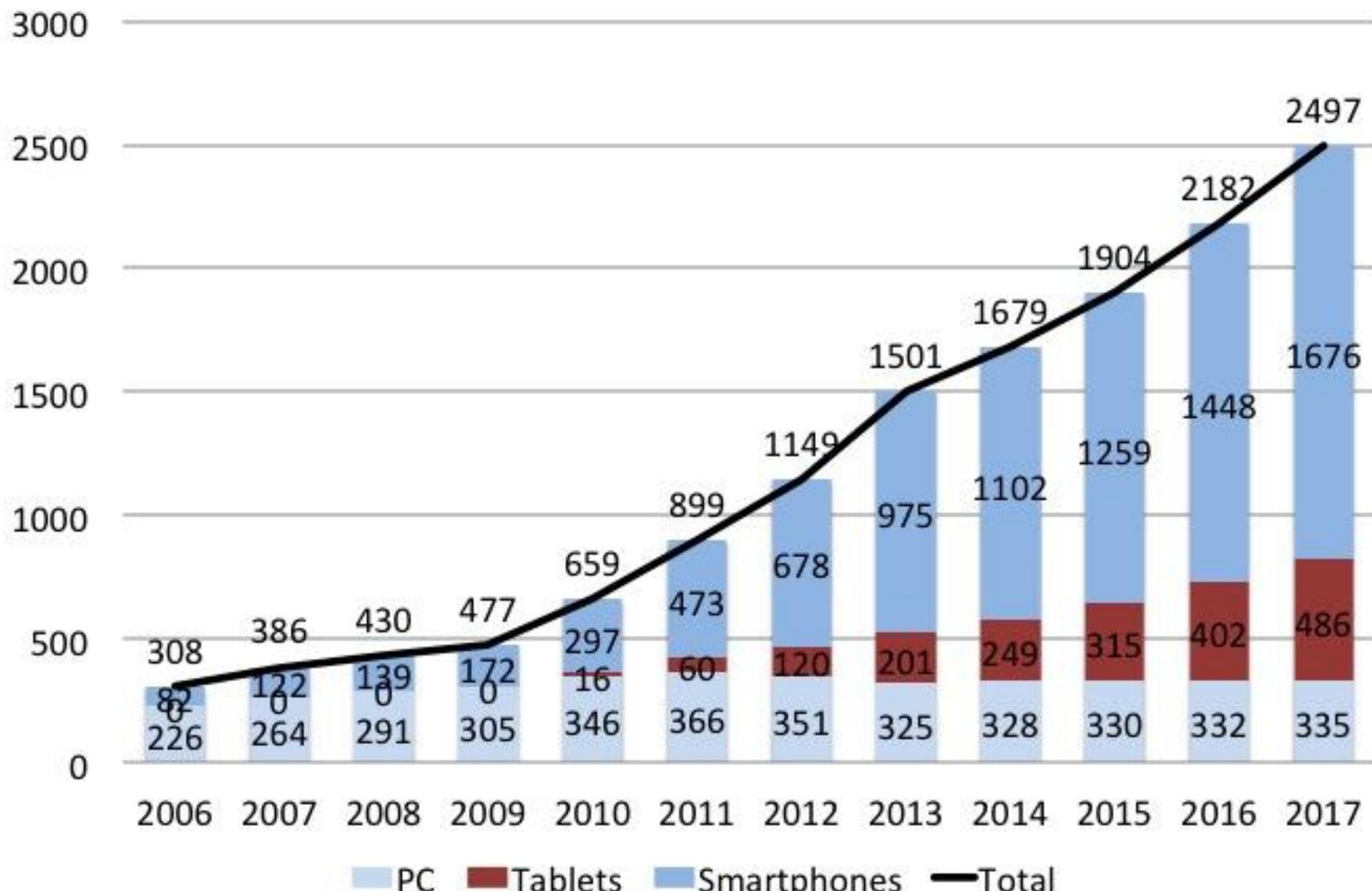
Enhance the Contrast?

## Device Shipment by Type



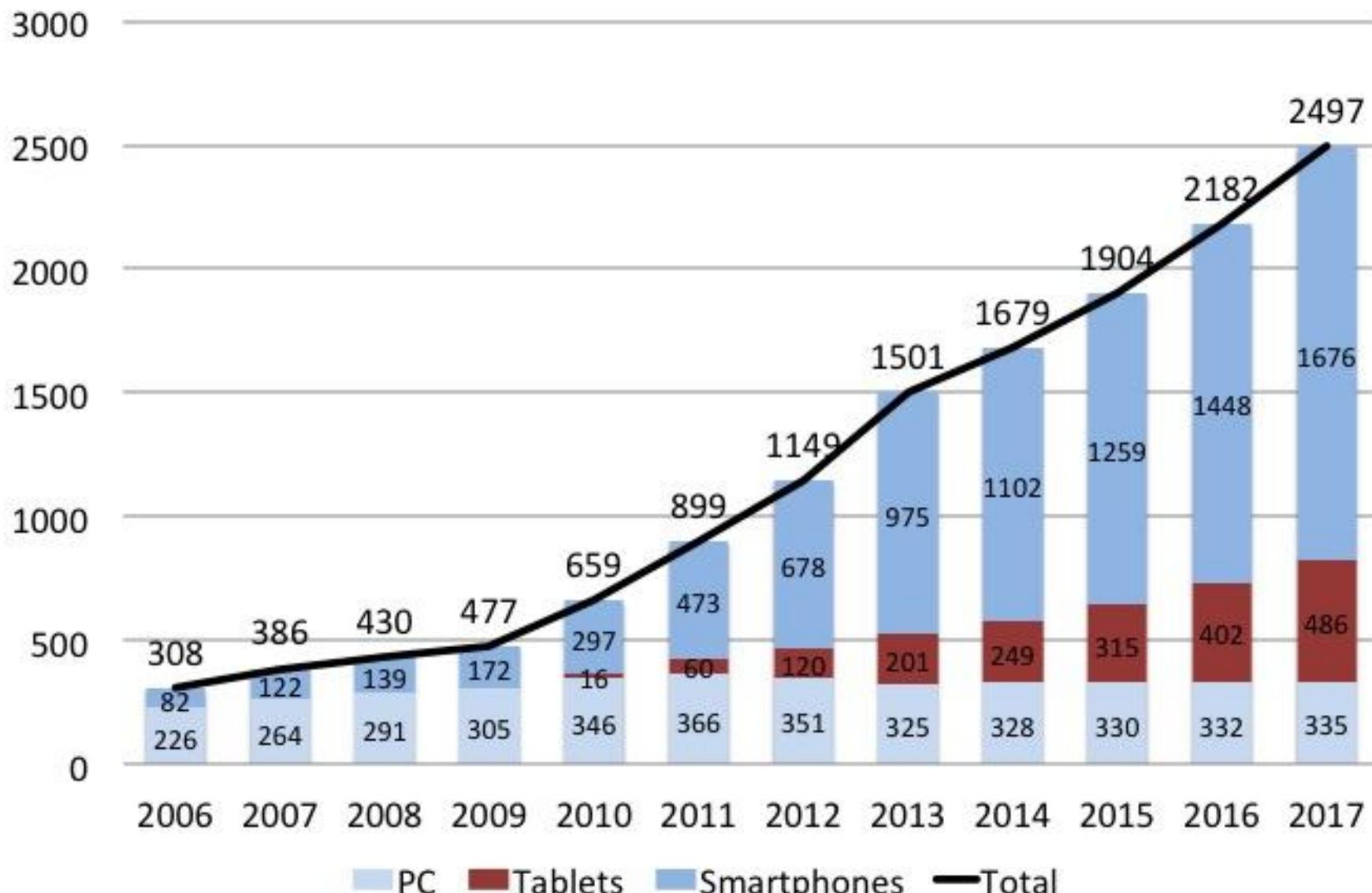
Does the Order Matter?

## Device Shipment by Type



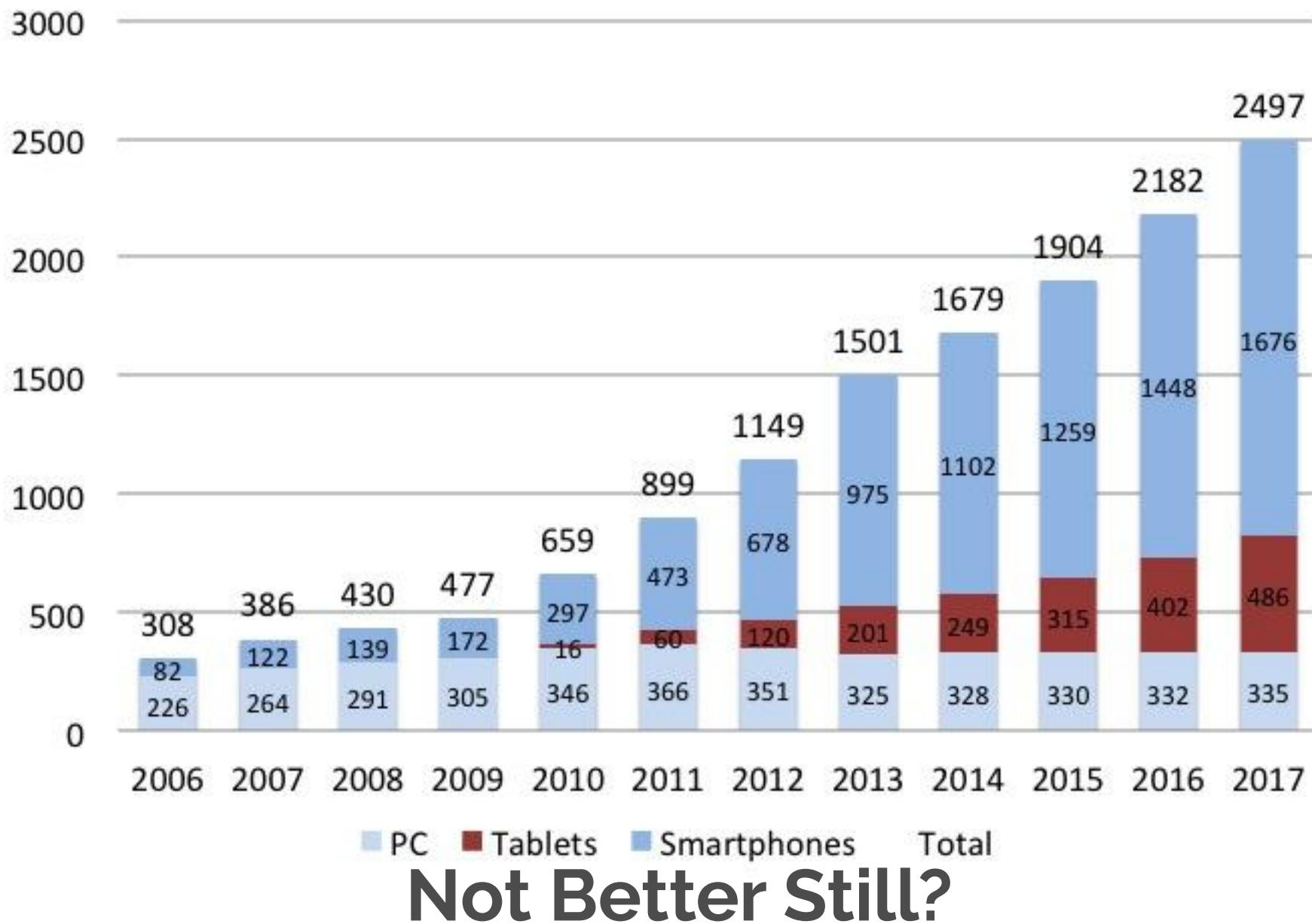
Add the Labels?

## Device Shipment by Type

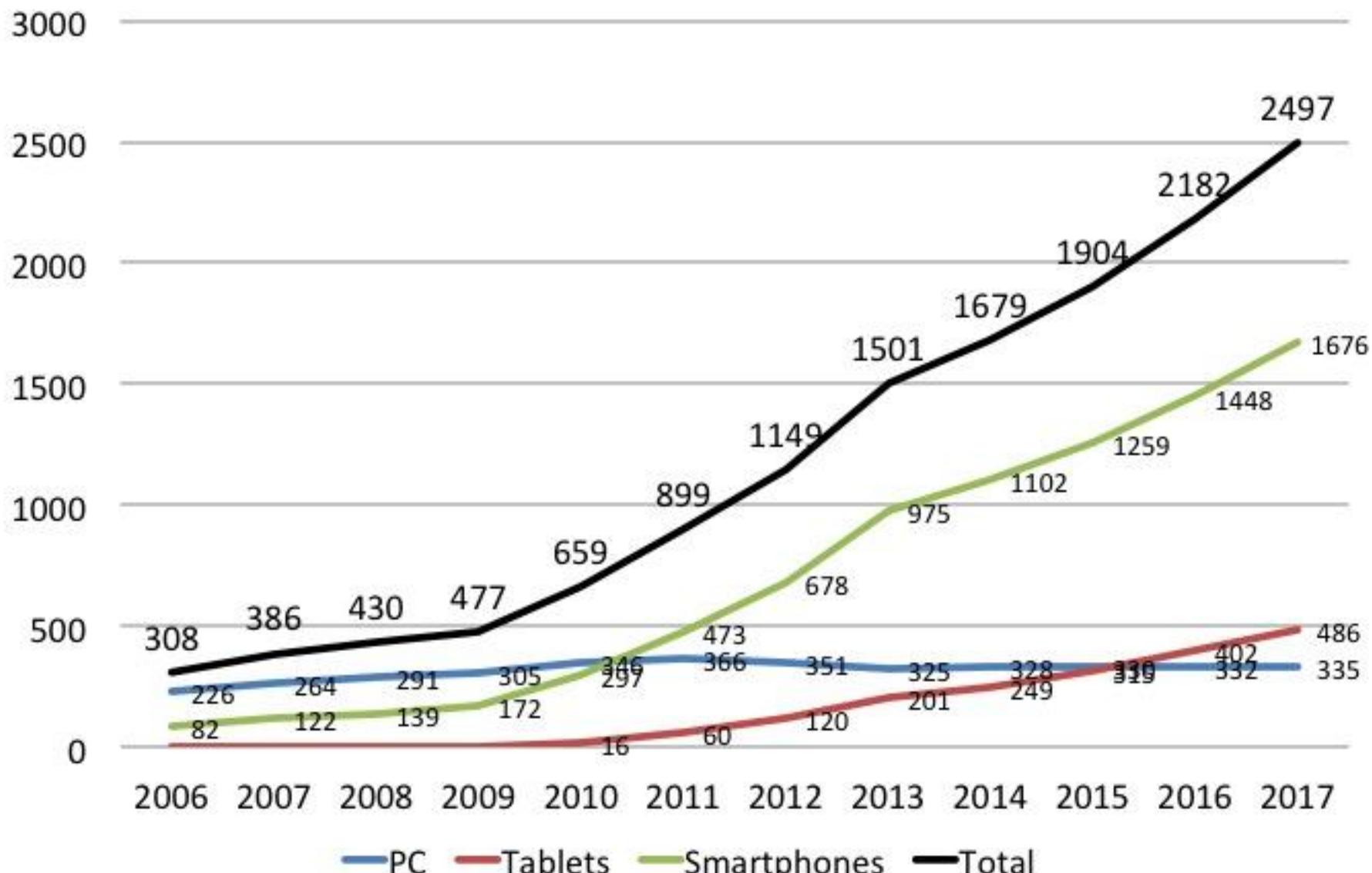


**Still too Busy?**

## Device Shipment by Type

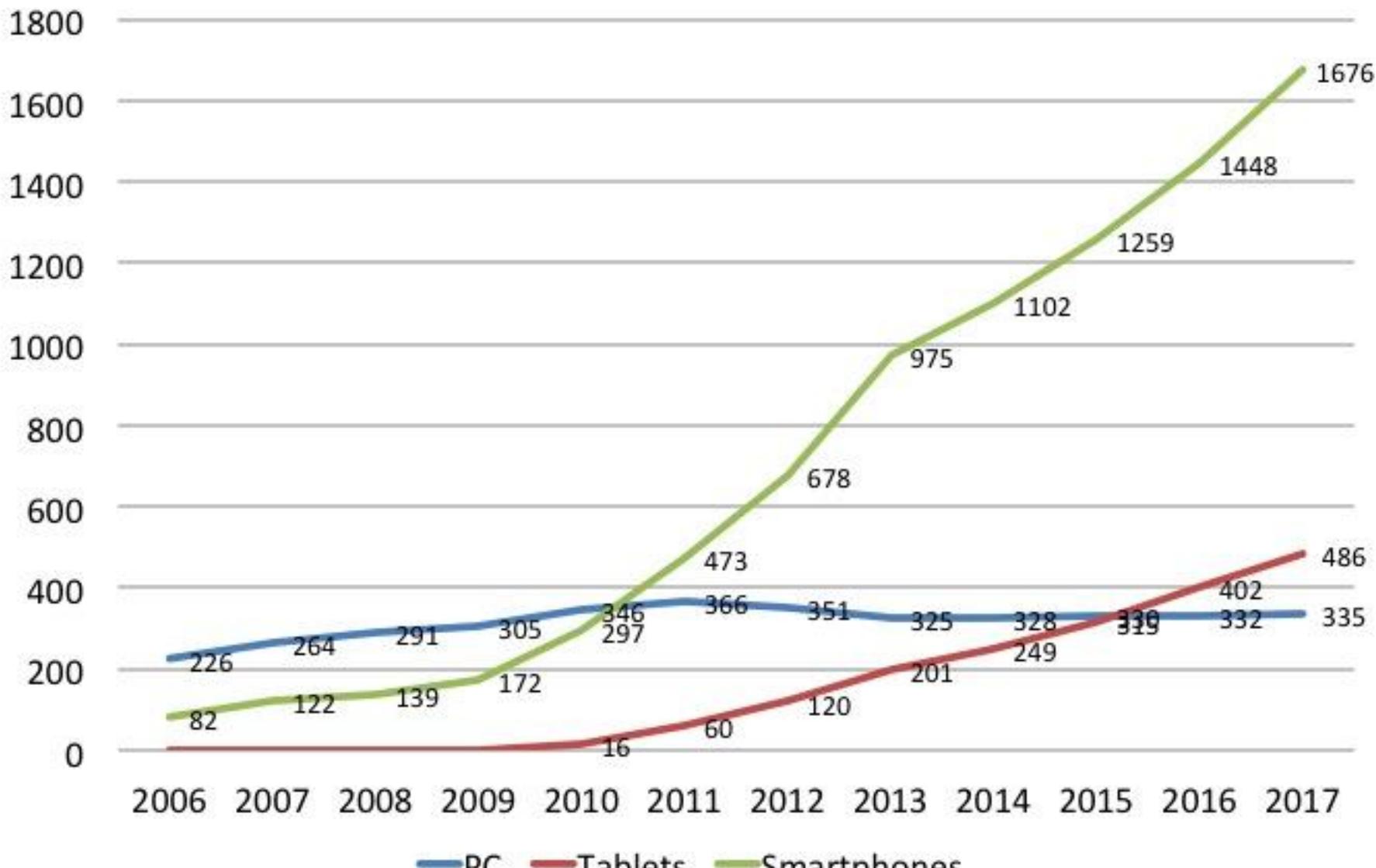


## Device Shipment by Type



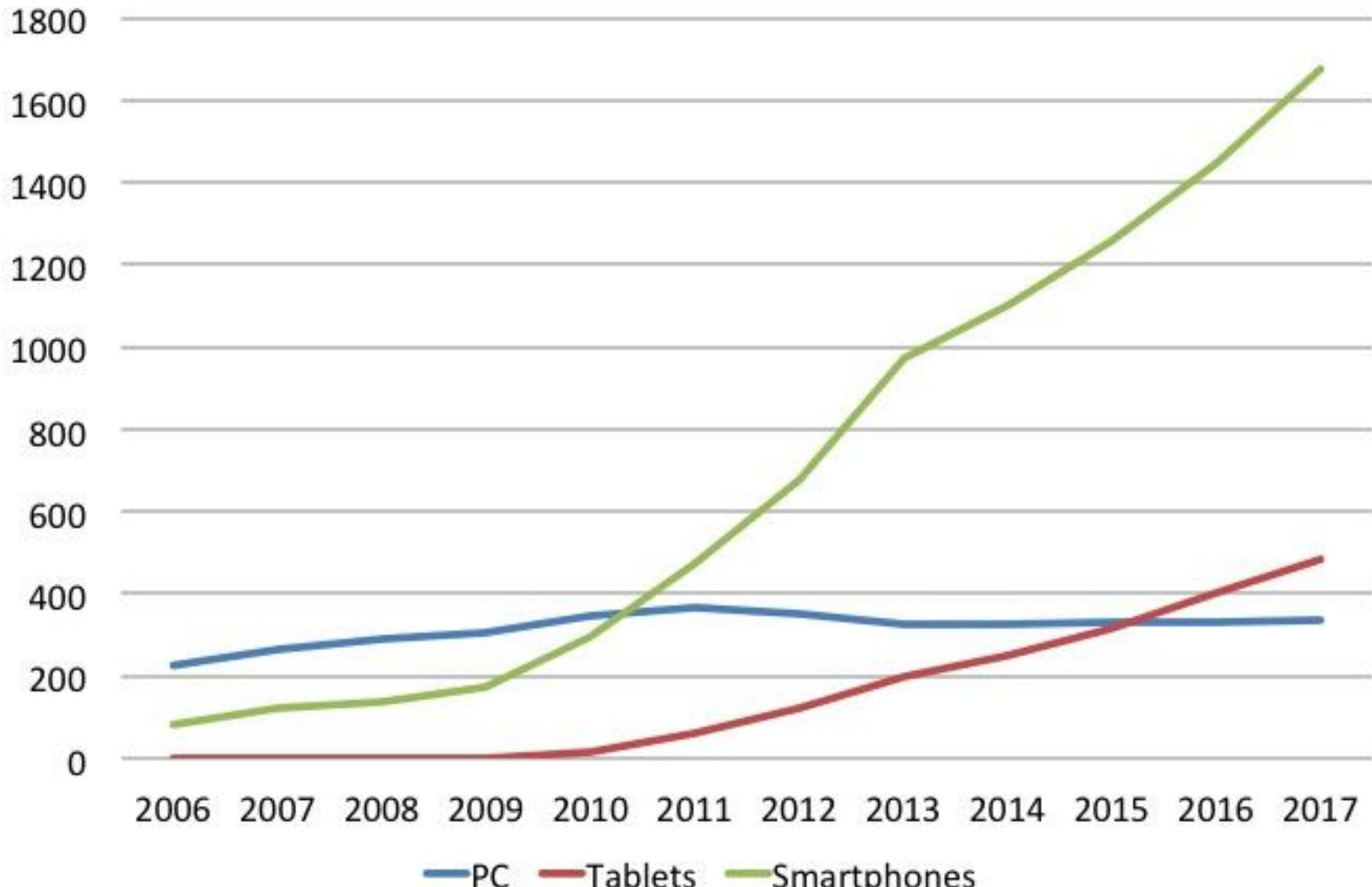
Show the trend?

## Device Shipment by Type



**Subtract more?**

## Device Shipment by Type



**Subtract even more?**

---

We are pretty good in data visualisation at handling the what, where, when questions, but we're bad at the why questions, which are often much more interesting.

Amanda Cox

---