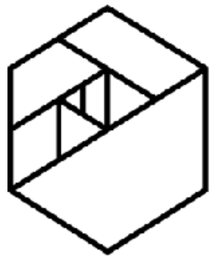
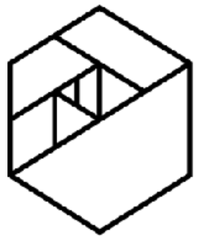


METIS

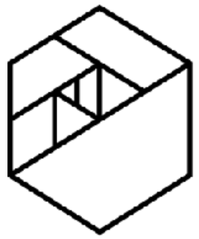
Day 4: Visualization

John Navarro
john.navarro@thisismetis.com
https://www.linkedin.com/in/johnnavarro/



METIS

Review Homework



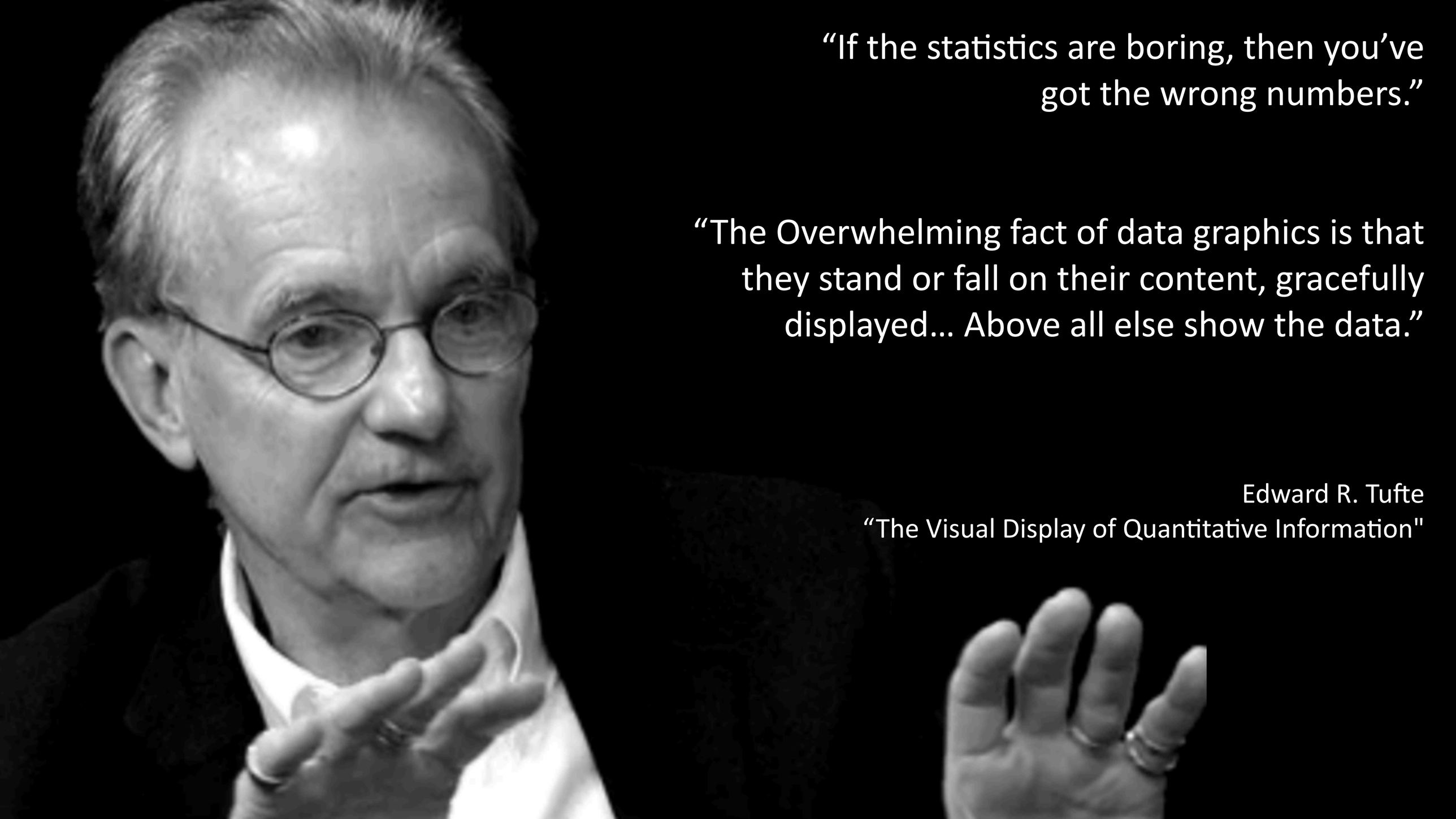
METIS

POP QUIZ!!

- **Scenario:** You found a million of what look like credit card numbers on a hidden website on the dark web.
- **Your task:** What metadata can you extract from these numbers? What kinds of analytic questions could
- HINT: You can read about credit card numbers here: **https://en.wikipedia.org/wiki/Payment_card_number**

Visualization Goals

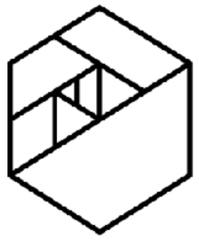
- Analyze
- Explore
- Assess
- Determine
- Decide
- Communicate
- Explain
- Present
- Prove
- Persuade



“If the statistics are boring, then you’ve got the wrong numbers.”

“The Overwhelming fact of data graphics is that they stand or fall on their content, gracefully displayed… Above all else show the data.”

Edward R. Tufte
“The Visual Display of Quantitative Information"



METIS

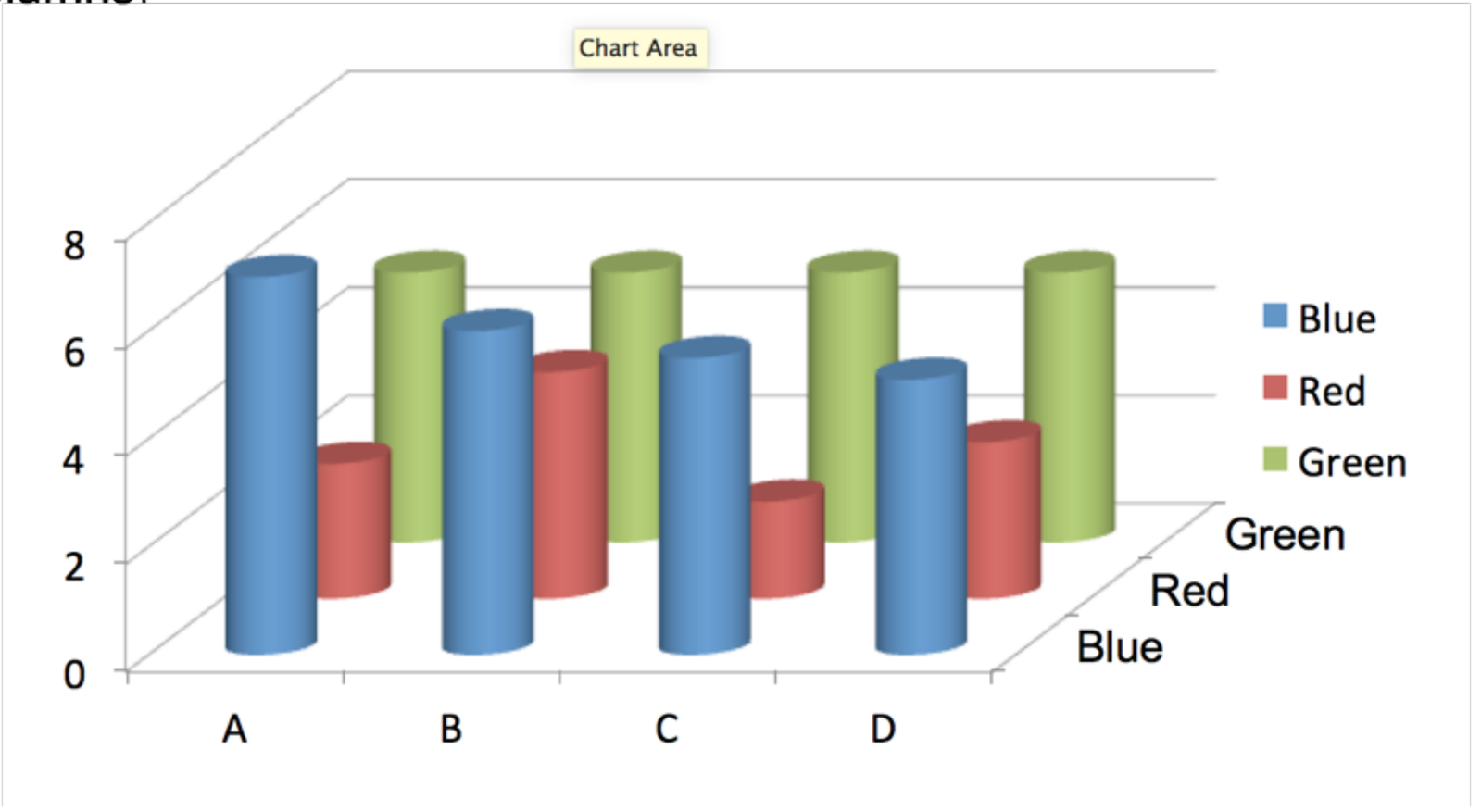
Elements of Good Visualizations

1. Graphical Integrity
2. Simple
3. Proper Display
4. Proper Color
5. Tells a story

METIS

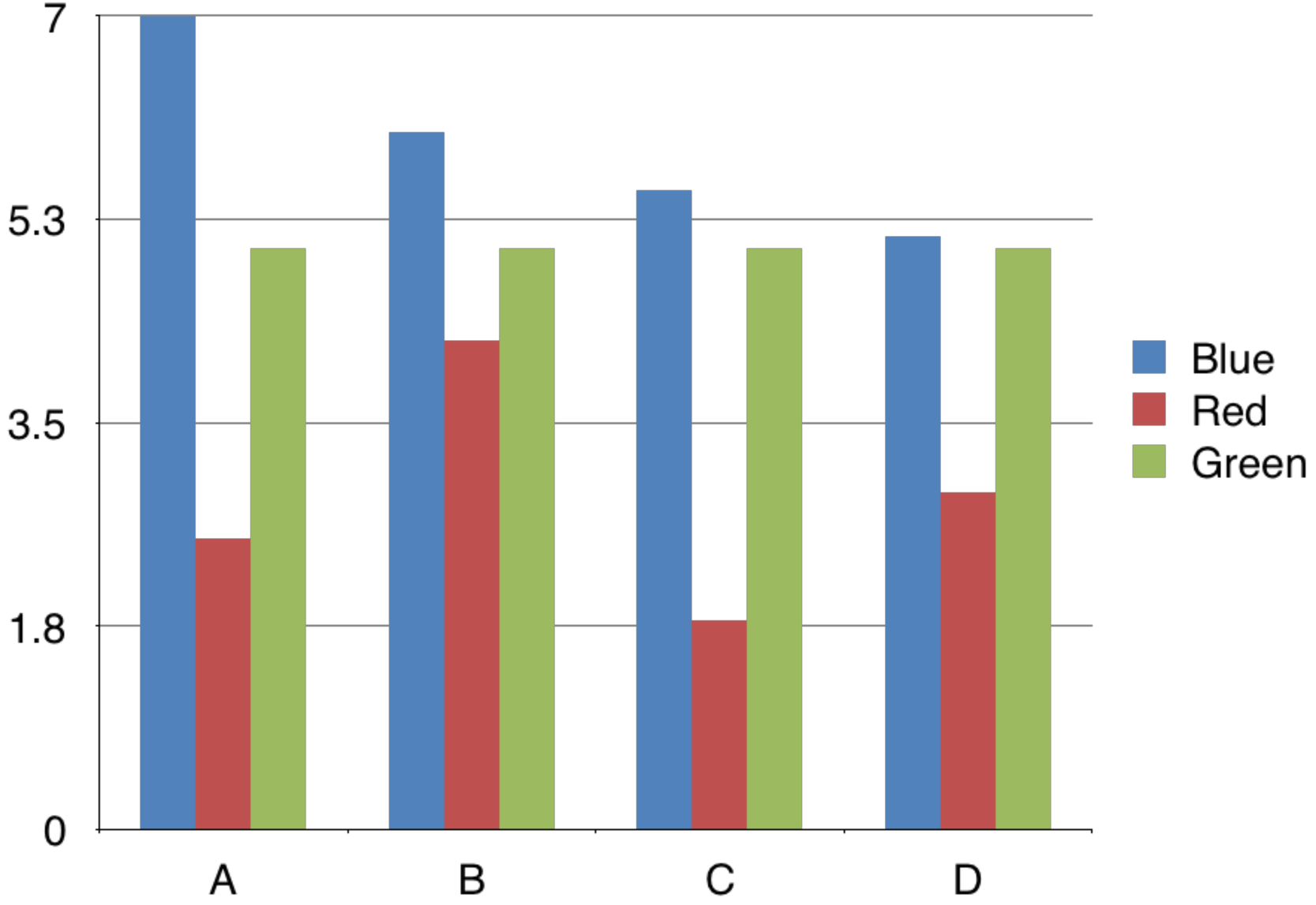
Proper Display

Questions: What is the height of the green columns? For which categories (A,B,C,D) are the blue columns taller than the green columns?



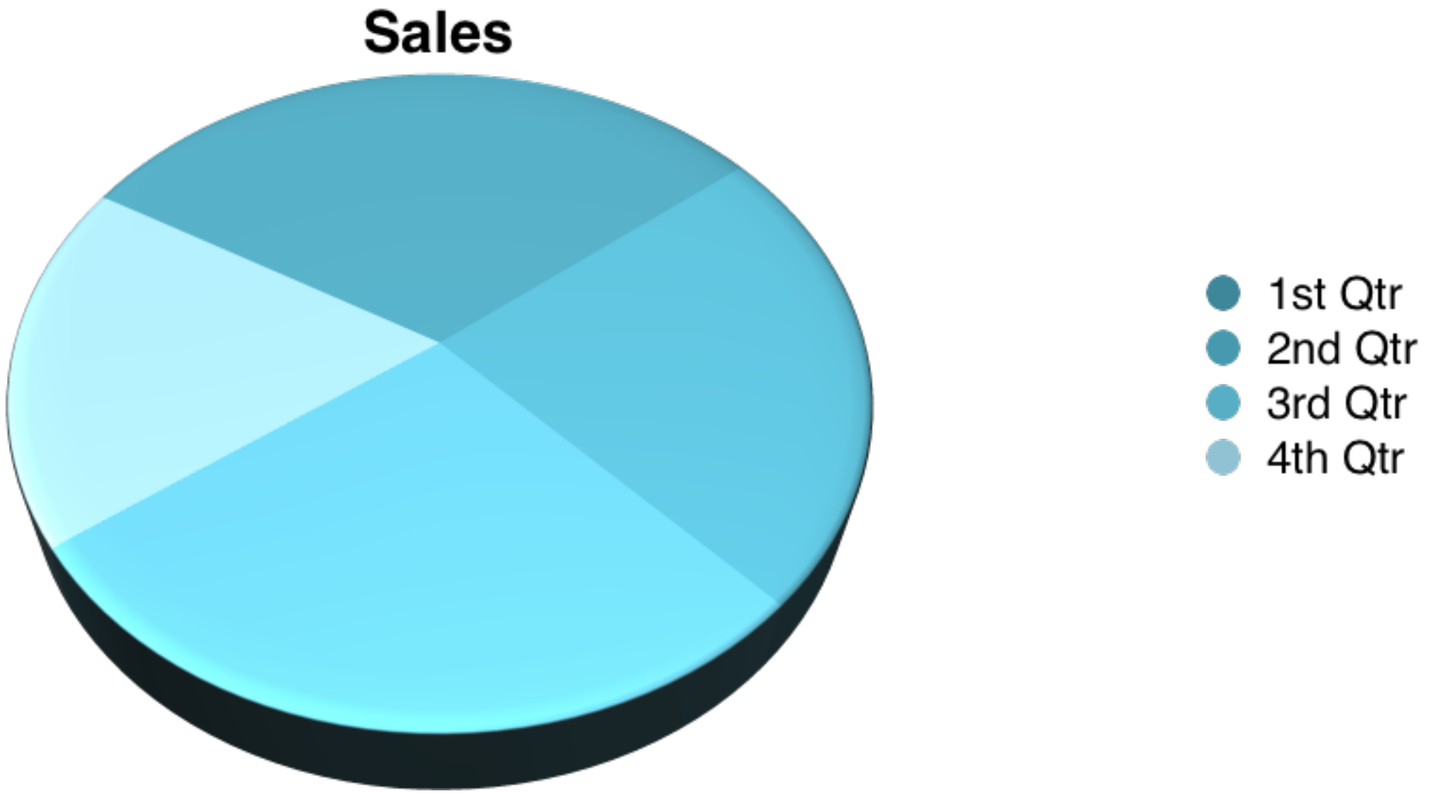
METIS

Proper Display



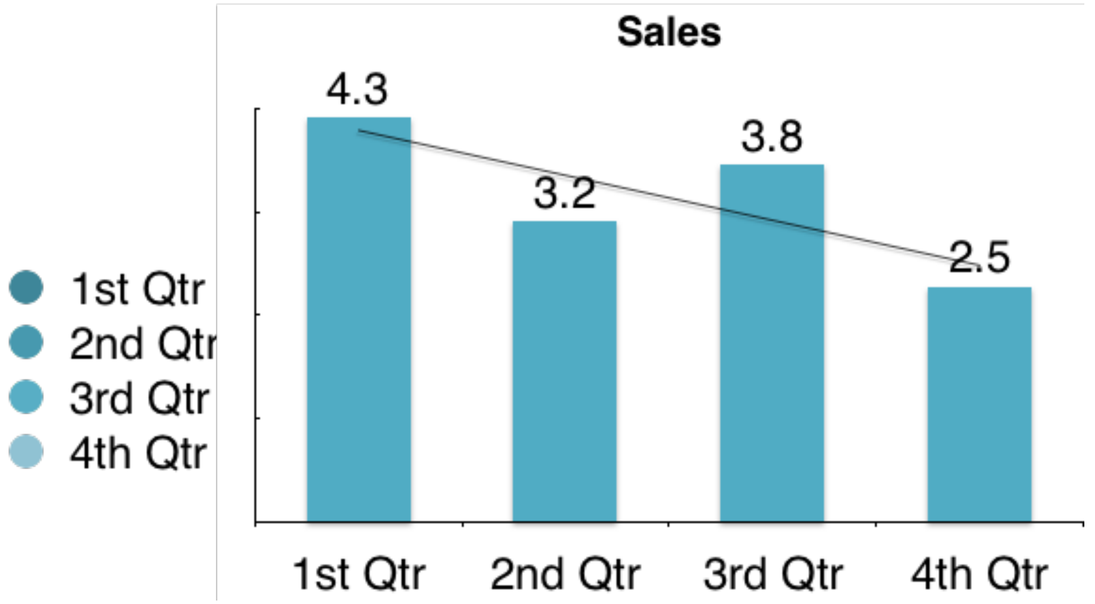
METIS

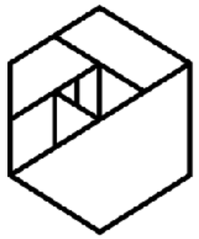
Proper Display



METIS

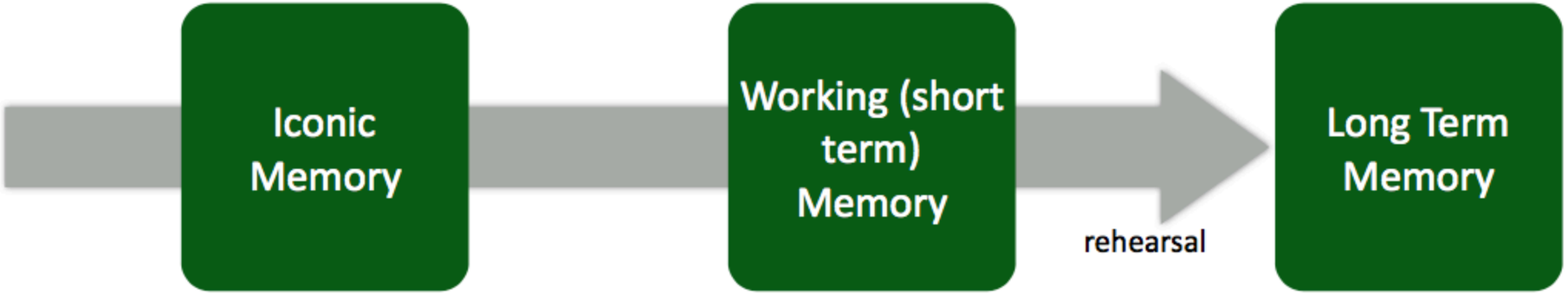
Proper Display





METIS

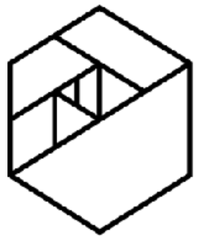
Visual Processing System



- < 1 second (or until replaced)
- Preattentive processing
- Saccadic movement
- ~20 objects

- Visual groupings
- Flexible, conscious
- 3-5 objects

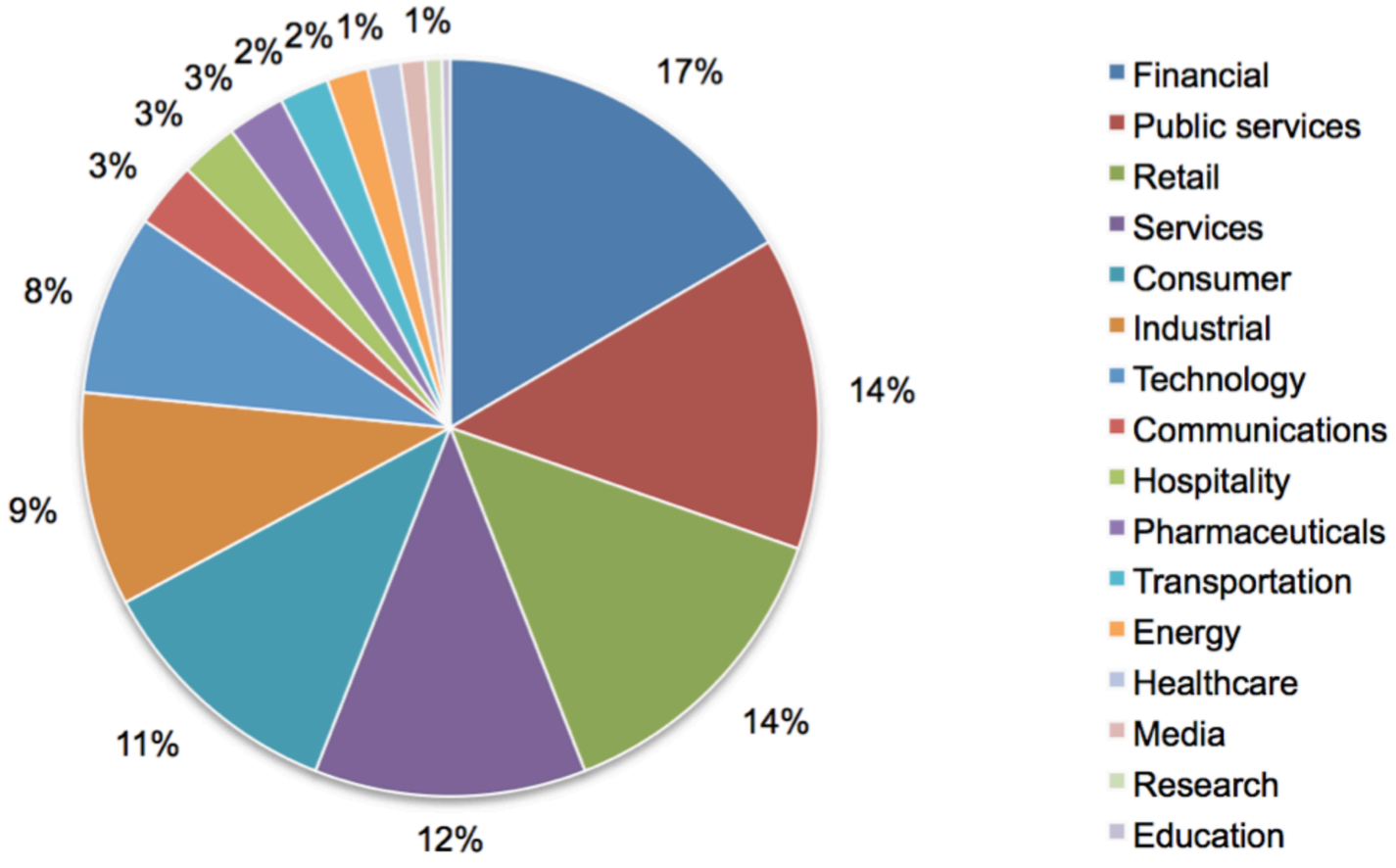
- Affects dataviz through expectations



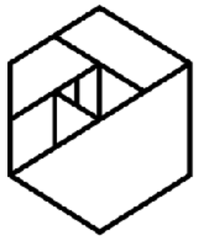
METIS

Overworking Visual Memory

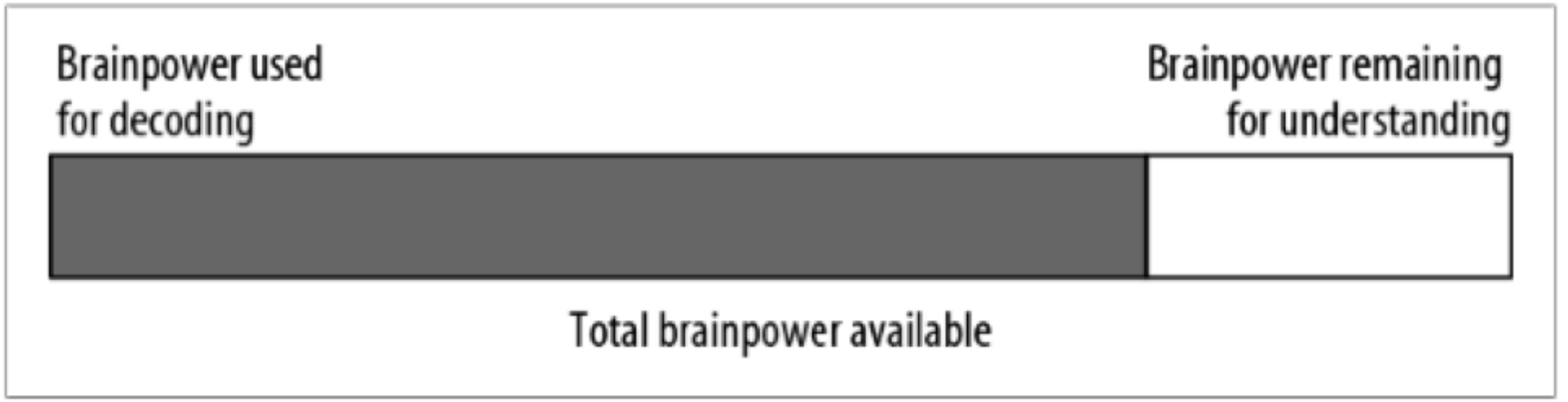
Figure 20. Distribution of the benchmark sample by industry segment
Consolidated (n = 277 organizations)

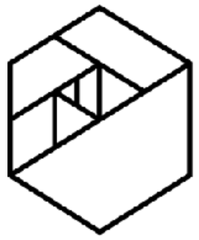


https://www4.symantec.com/mktginfo/whitepaper/053013_GL_NA_WP_Ponemon-2013-Cost-of-a-Data-Breach-Report_daiNA_cta72382.pdf



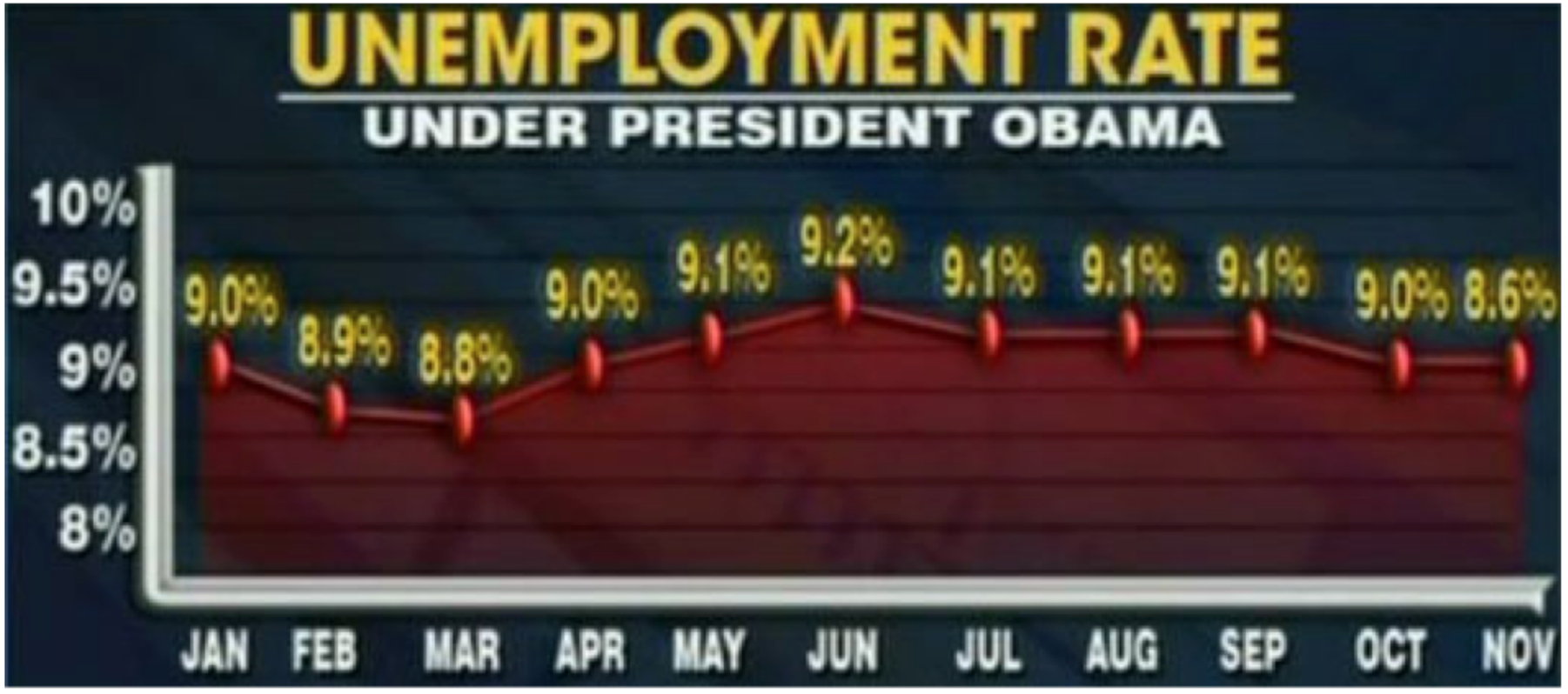
METIS


Brainpower used
for decoding
Brainpower remaining
for understanding
Total brainpower available

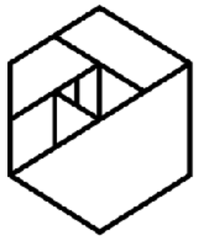


METIS

Graphical Integrity

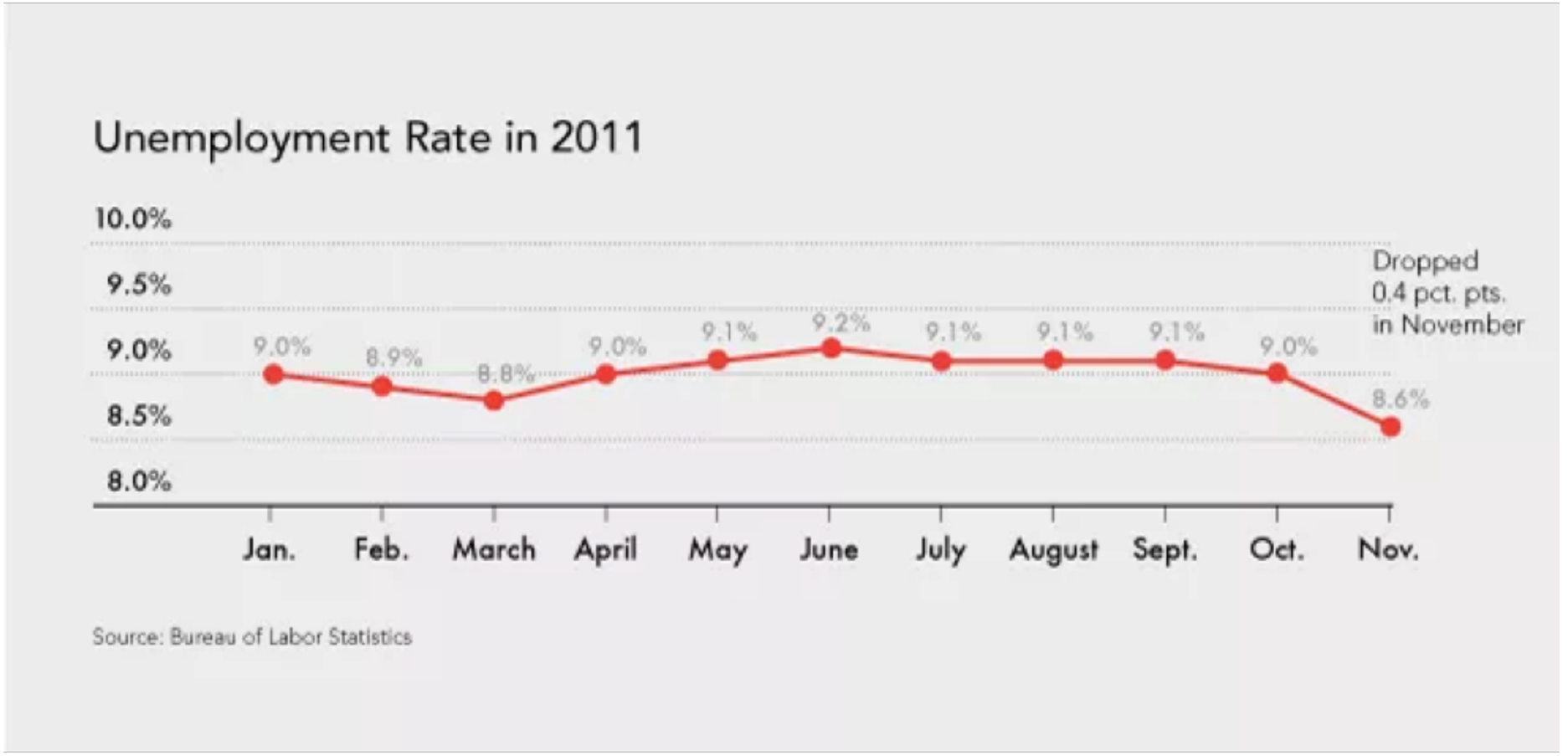


flowingdata.com



METIS

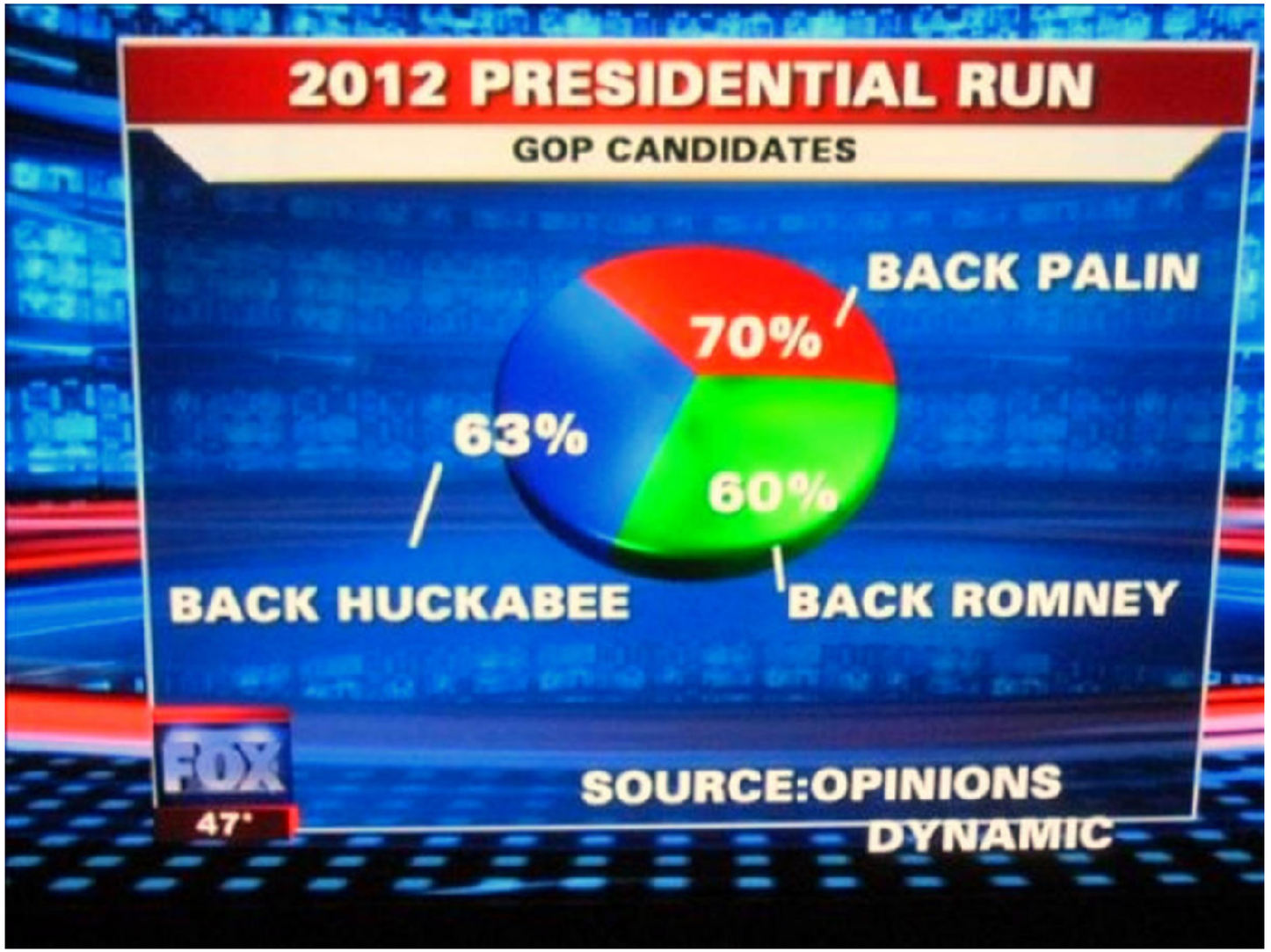
Graphical Integrity

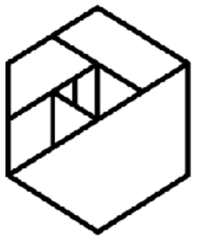


flowingdata.com

METIS

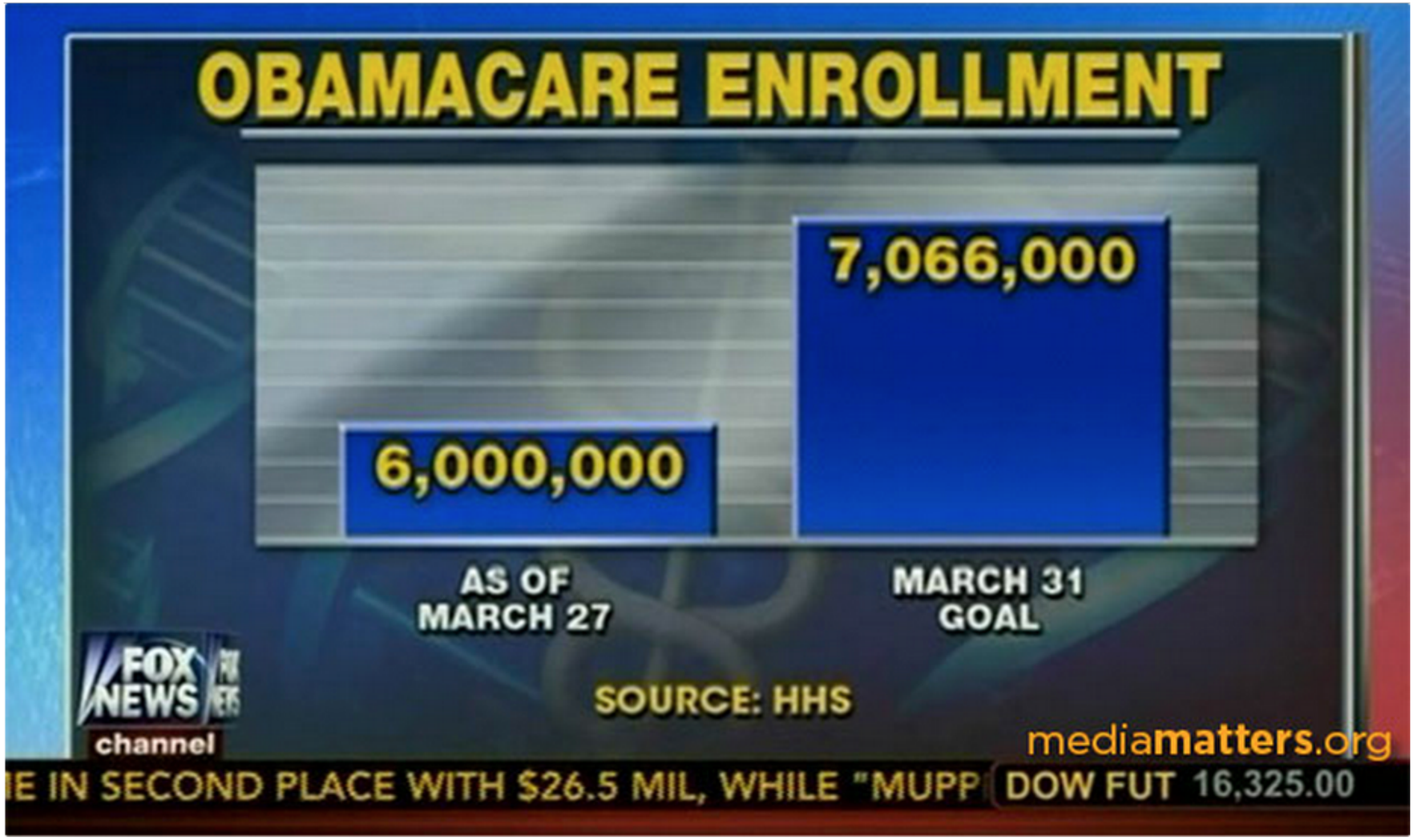
Graphical Integrity?





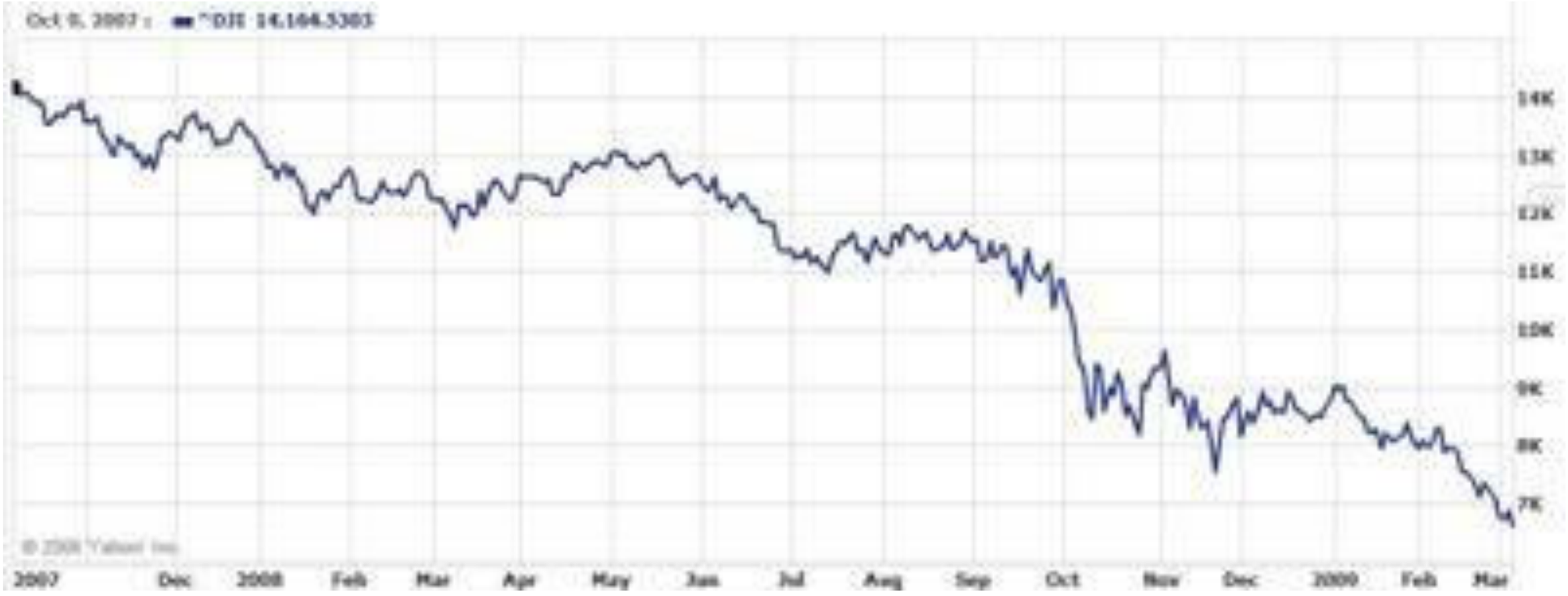
METIS

Graphical Integrity



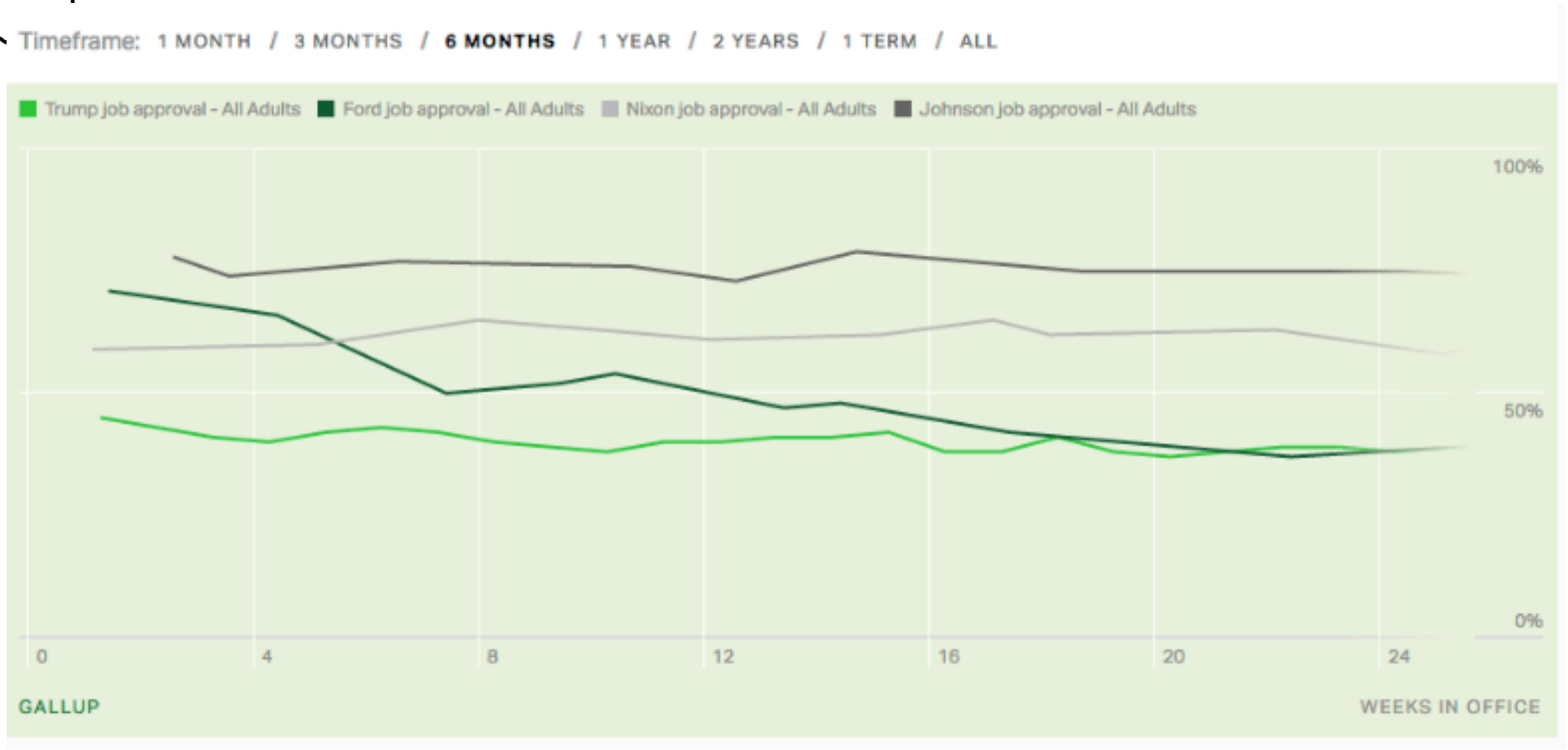

STOCK MARKET SLIDE
NOV. 4, 2008
ELECTION DAY
9625.30
JAN. 20, 2009
INAUGURATION DAY
7949.10
FRIDAY
MAR. 5, 2009
6626.94
SOURCE: BUSINESSWEEK
6:31 CT
MSNBC

METIS


Oct 9, 2007
^DJI 14,164.5303
14K
13K
12K
11K
10K
9K
8K
7K
2007
Dec
2008
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2009
Feb
Mar


GEORGIA
PRESIDENT
FEELINGS ABOUT THE FEDERAL GOVERNMENT
SATISFIED
D CLINTON 81%
R TRUMP 14%
JOHNSON 4%
CNN EXIT POLL
26:23
NEXT POLLS CLOSE
LIVE
CNN
CNN ELECTORAL MAP
19 TRUMP
3 CLINTON
IN PRESIDENT
R TRUMP 126,791 65.4%
D CLINTON 59,284 30.6%

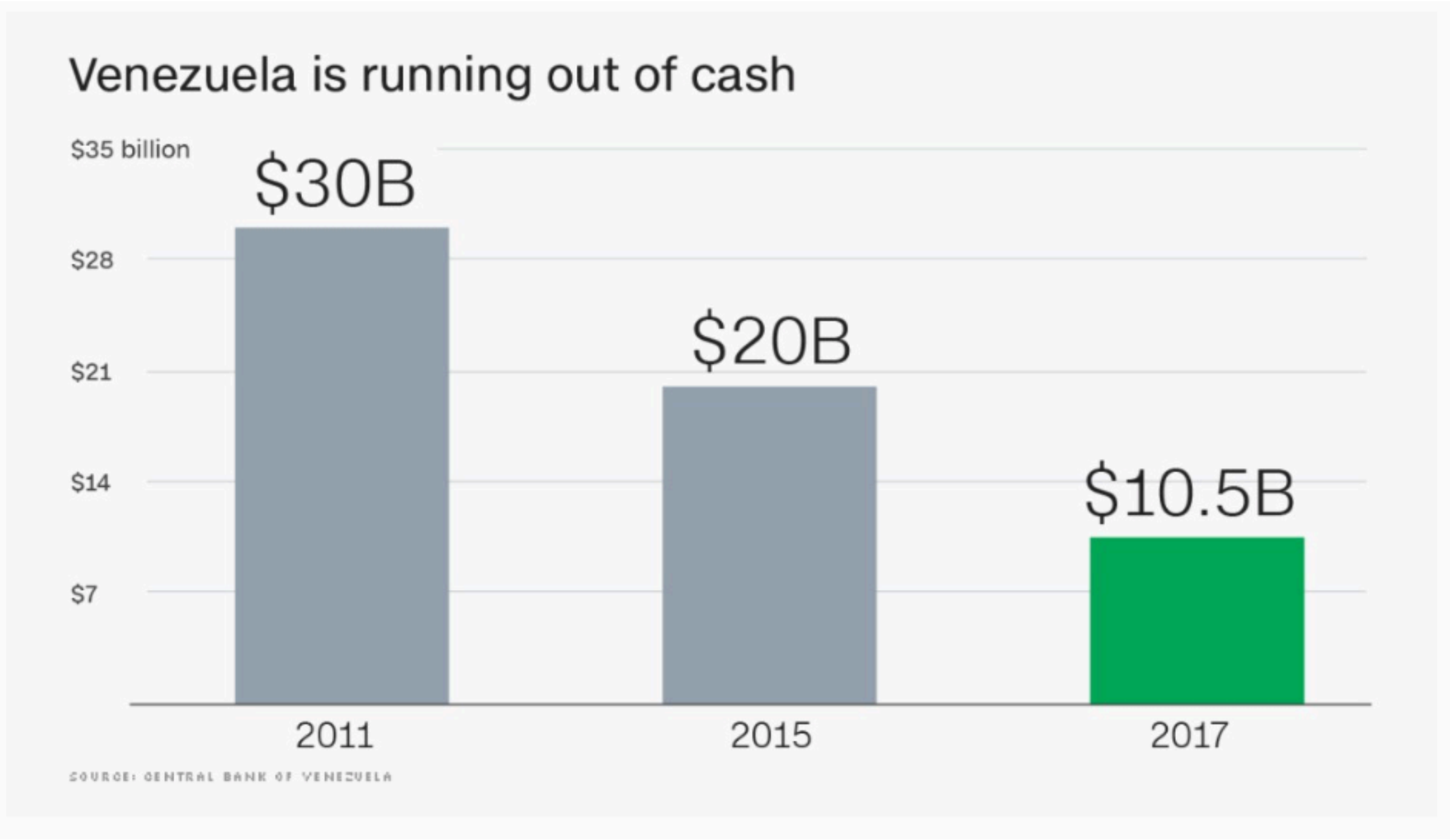
METIS



http://www.cnn.com/2017/07/17/politics/trump-approval-ratings/index.html

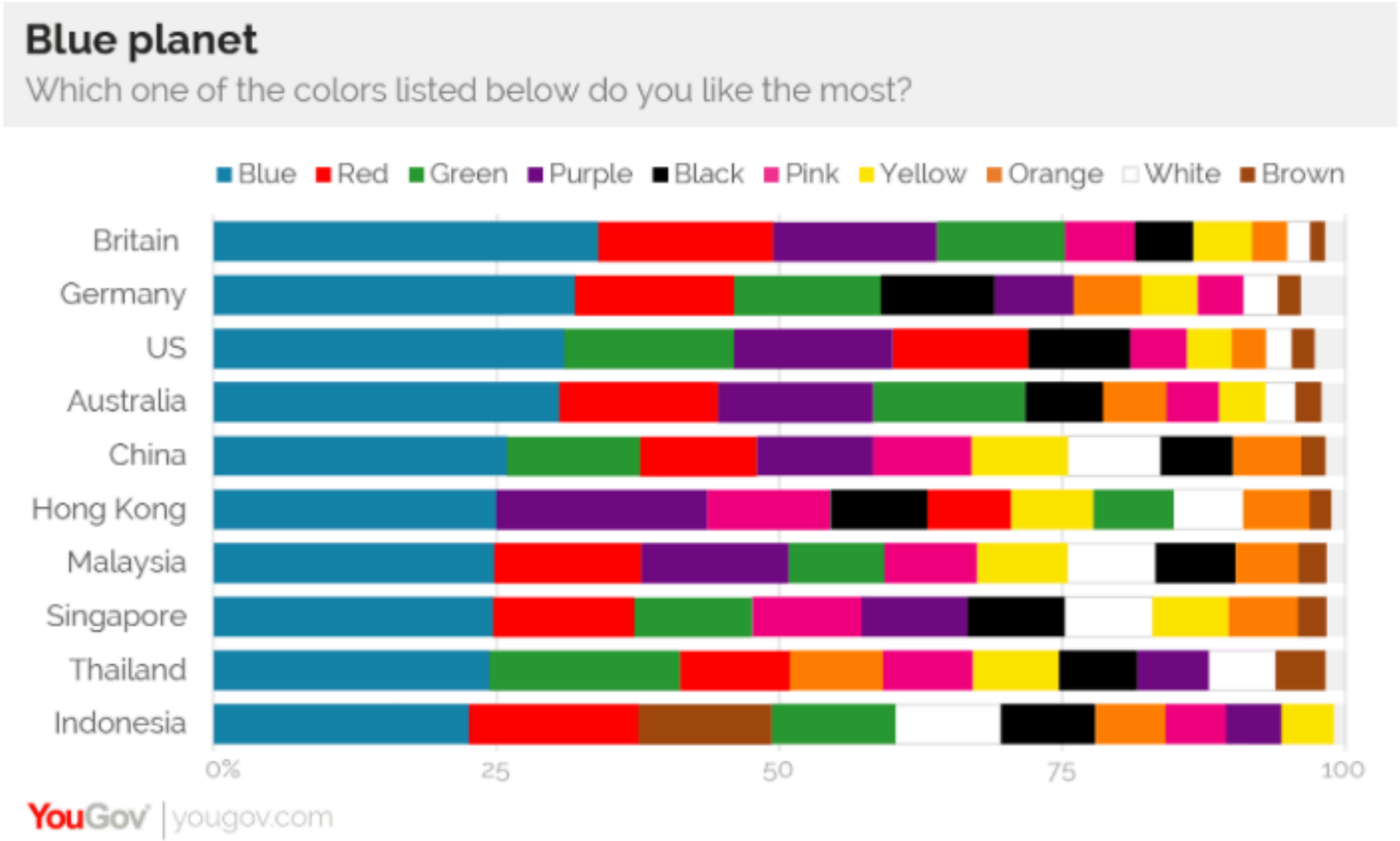
METIS

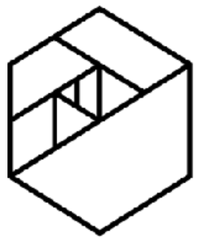
Graphical Integrity



METIS

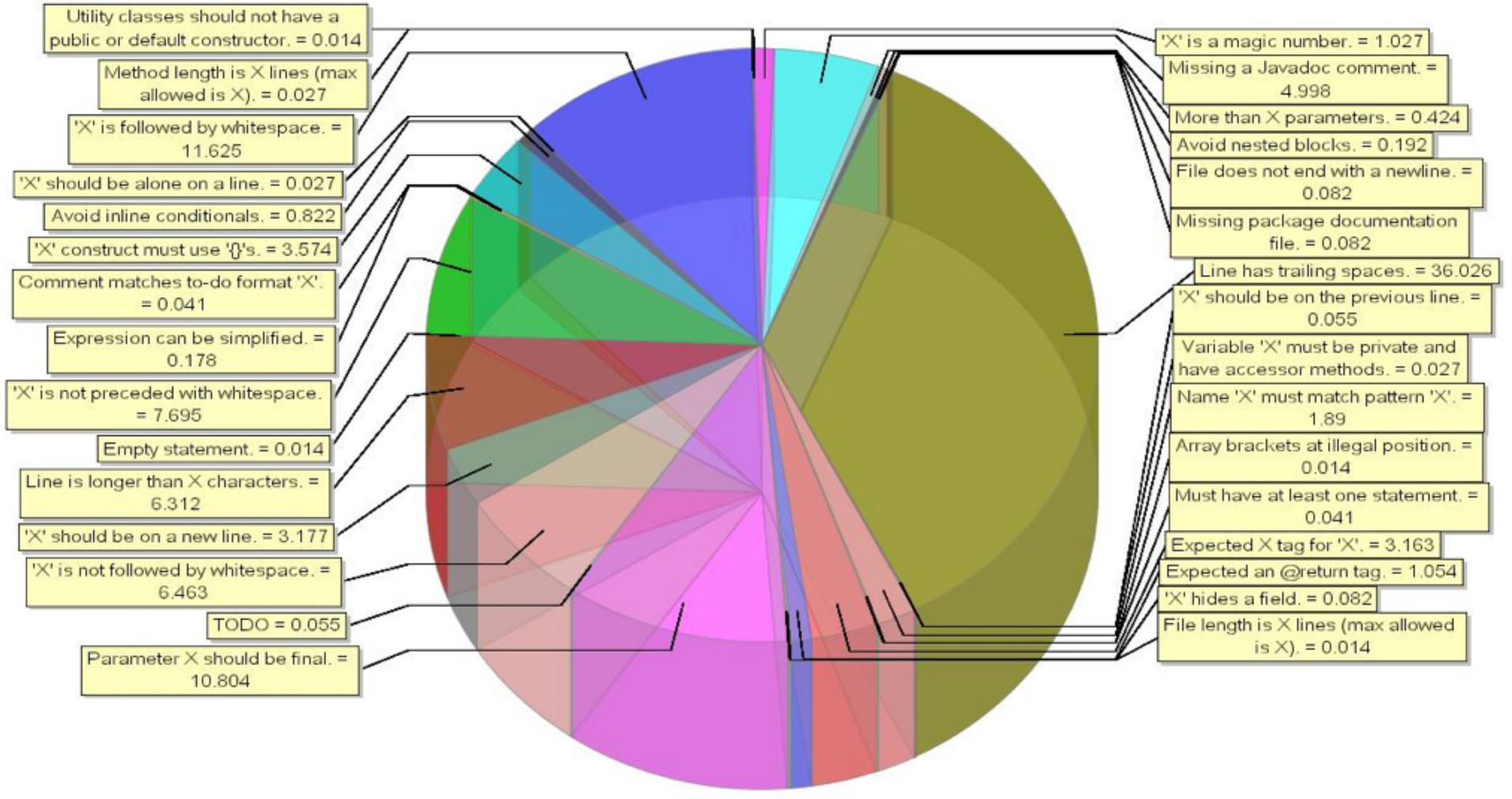
Graphical Integrity





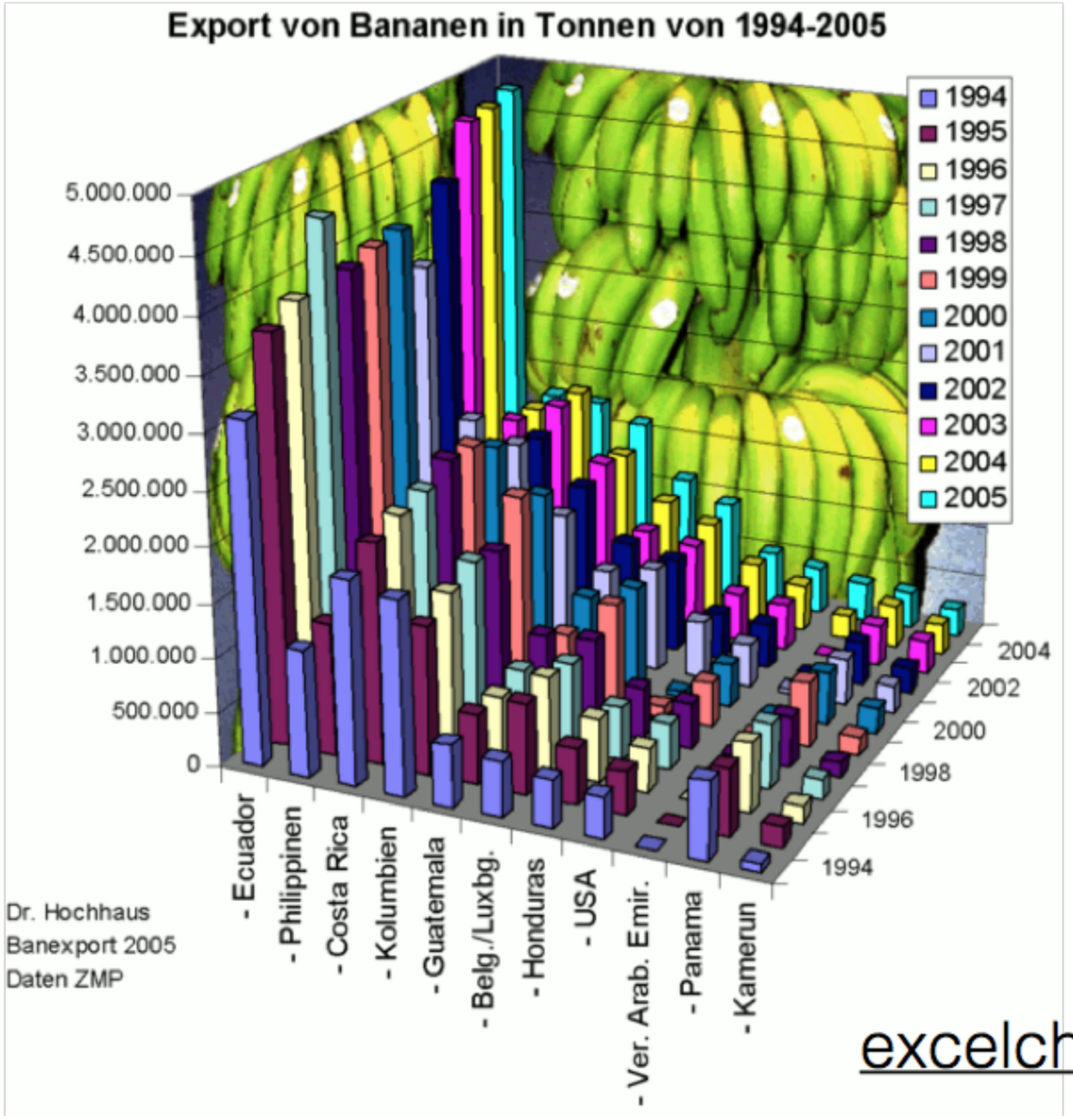
METIS

Simple

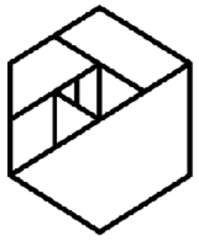


METIS

Simple



excelcharts.com



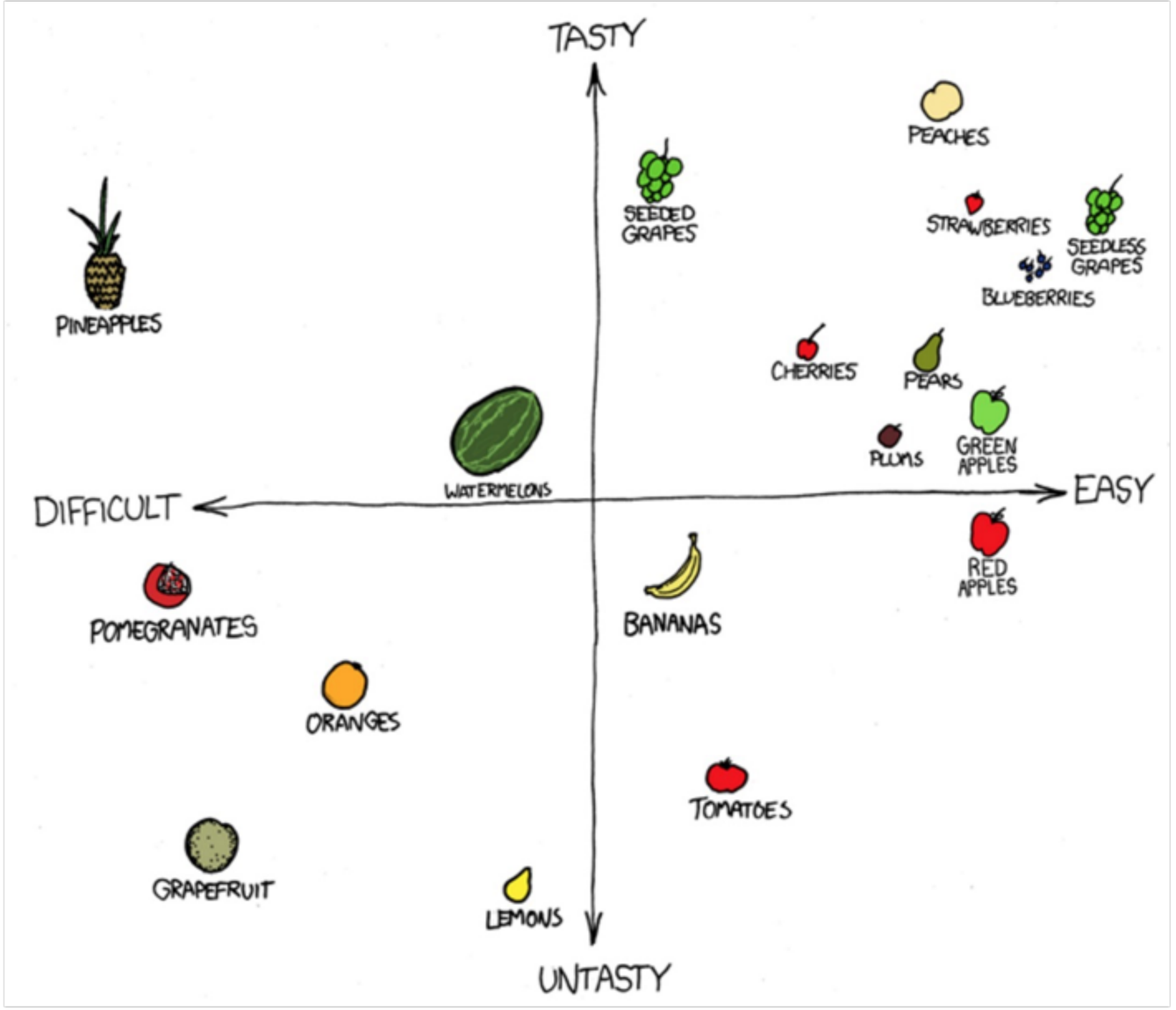
METIS

Proper Display

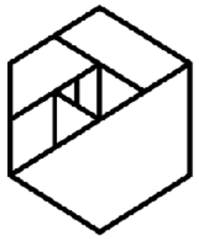
Google Finance

METIS

Proper Display



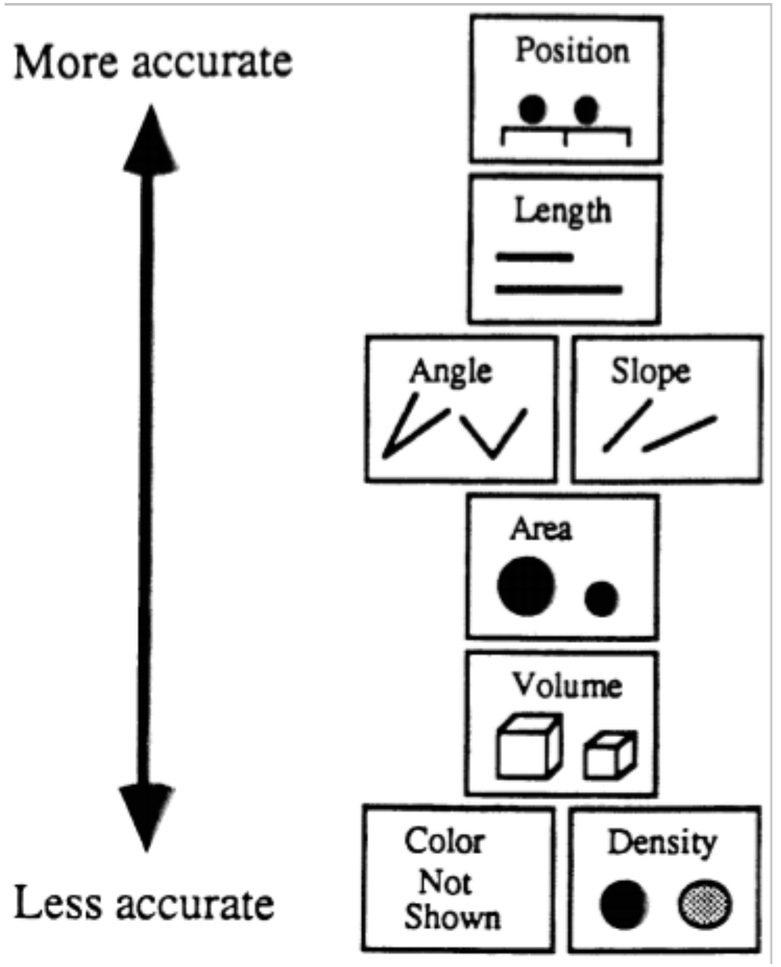
xkcd.com/388

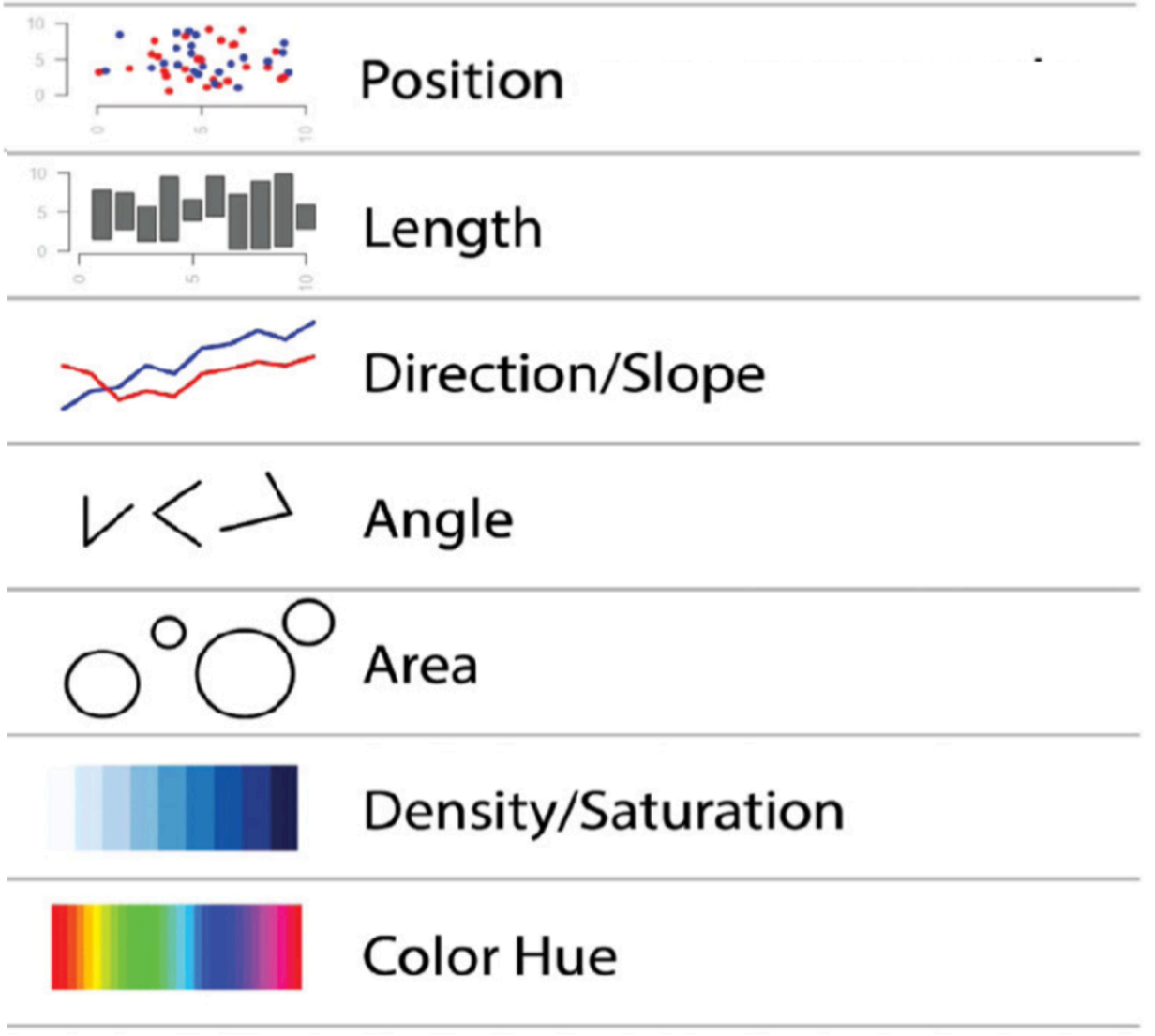


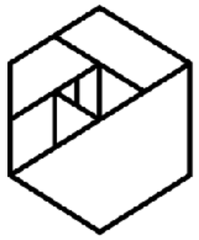
METIS

Visual Encoding

METIS


More accurate
Position
Length
Angle
Slope
Area
Volume
Color Not Shown
Density
Less accurate

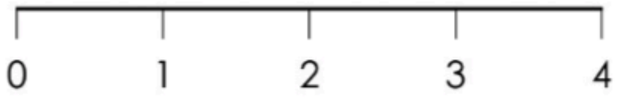

Position
Length
Direction/Slope
Angle
Area
Density/Saturation
Color Hue



METIS

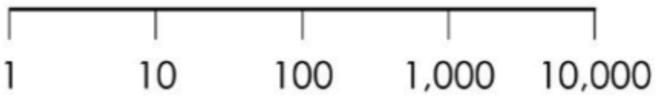
Linear

Values are evenly spaced



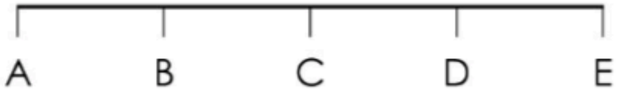
Logarithmic

Focus on percent change



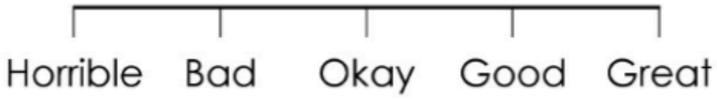
Categorical

Discrete placement in bins



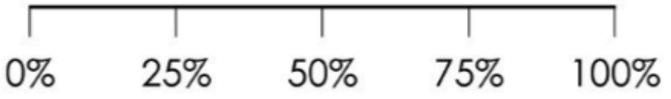
Ordinal

Categories where order matters



Percent

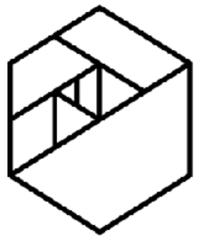
Representing parts of a whole



Time

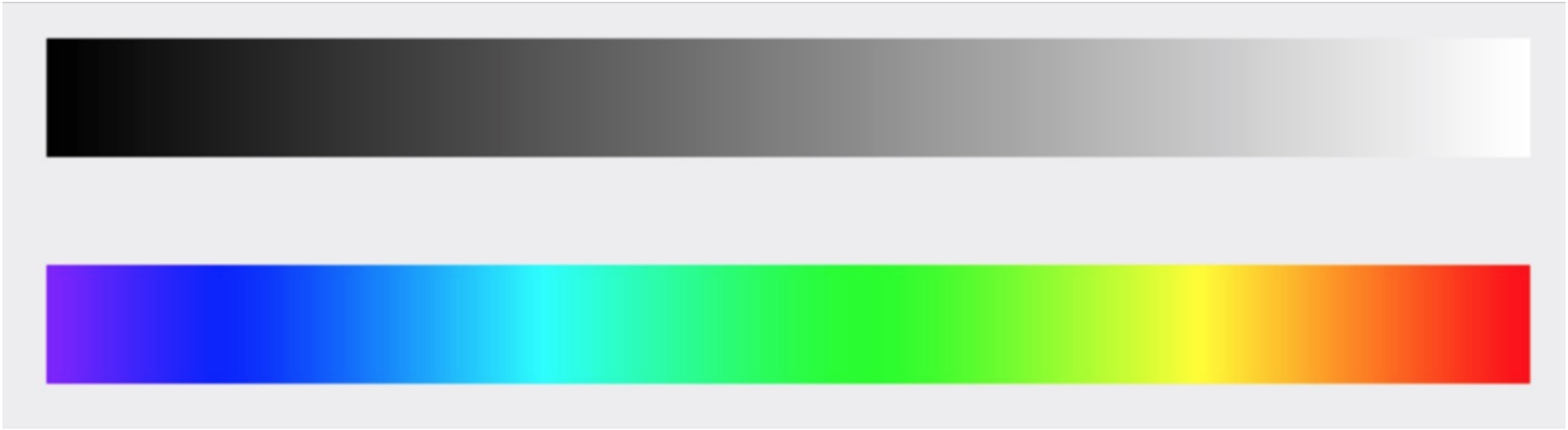
Units of months, days, or hours

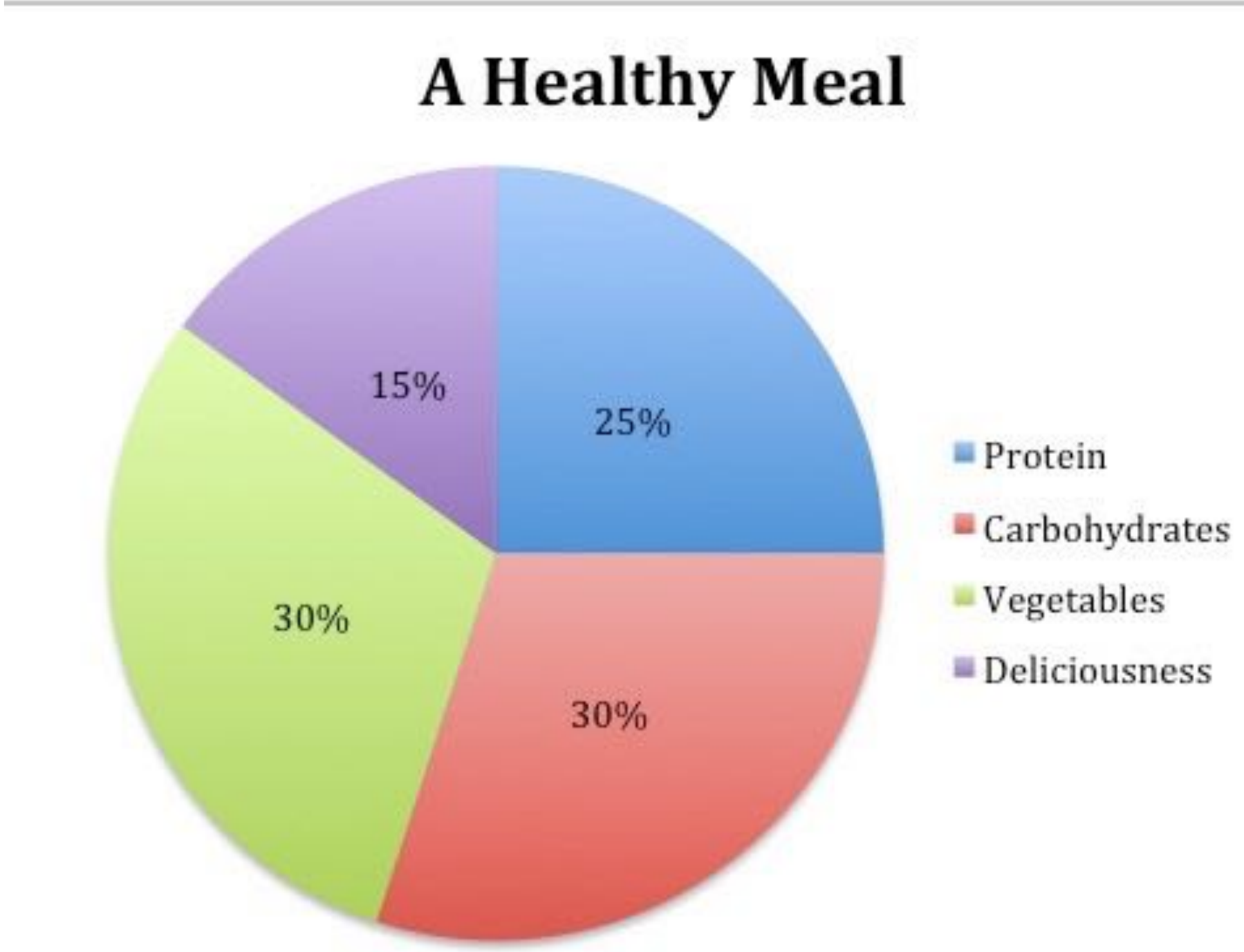
Source: Nathan Yau, Data Points

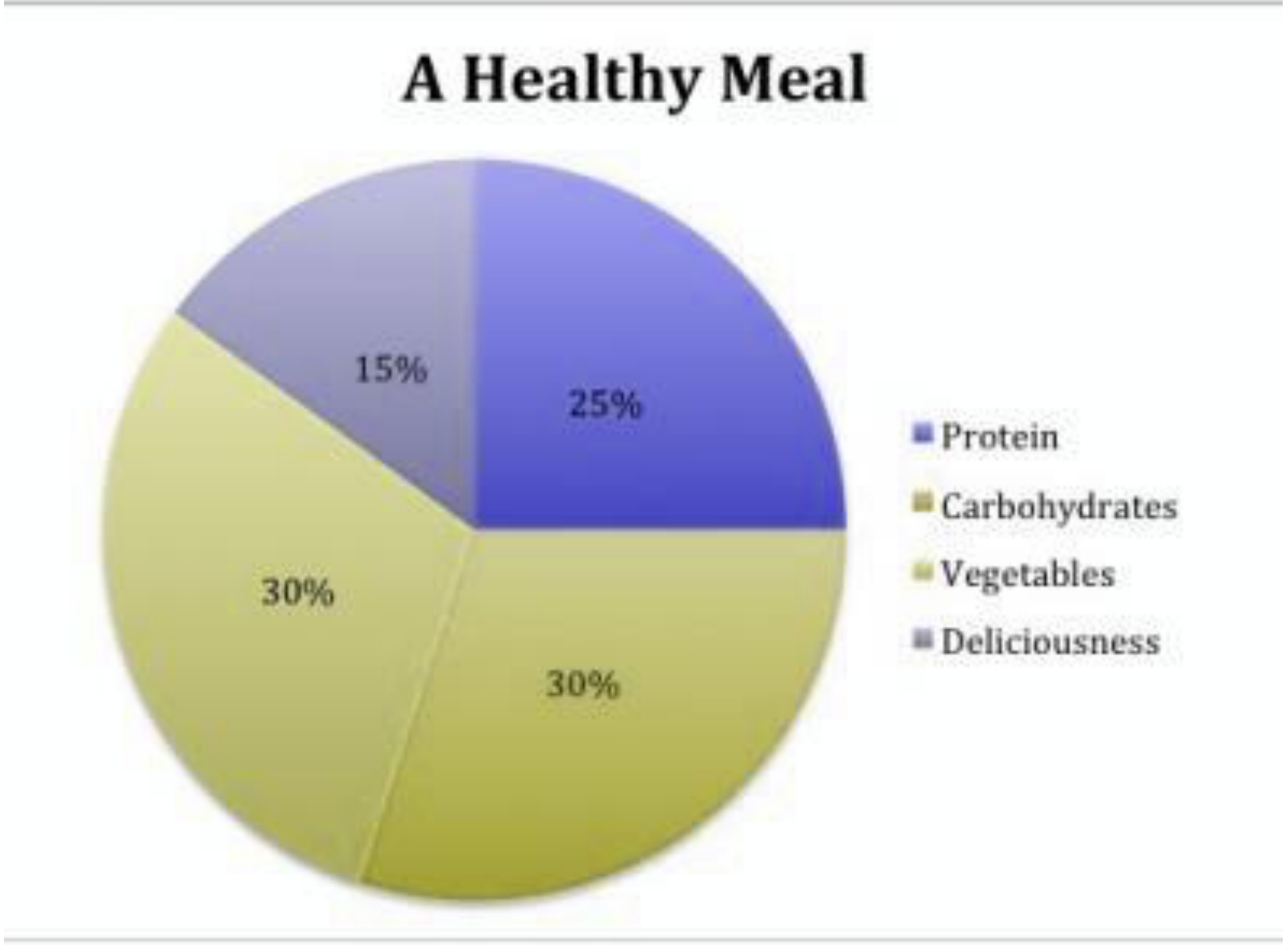


METIS

Color




A Healthy Meal
25%
30%
30%
15%
Protein
Carbohydrates
Vegetables
Deliciousness


A Healthy Meal
15%
25%
30%
30%
Protein
Carbohydrates
Vegetables
Deliciousness


elastica
Dashboard
Users
Total (192)
28
High Risk
0
Med Risk
4
Blocked
Policies
Total (231)
3
Blocking
142
Alerting
Policy Alerts
Alerting
443
Blocked
13
Rest
0
Threat Alerts
High Risk
2
Med Risk
2
Low Risk
17
Audited Services
by Users
High Risk (736)
Medium Risk (3k)
Low Risk (3k)
3k
494.3 GB
963k
243

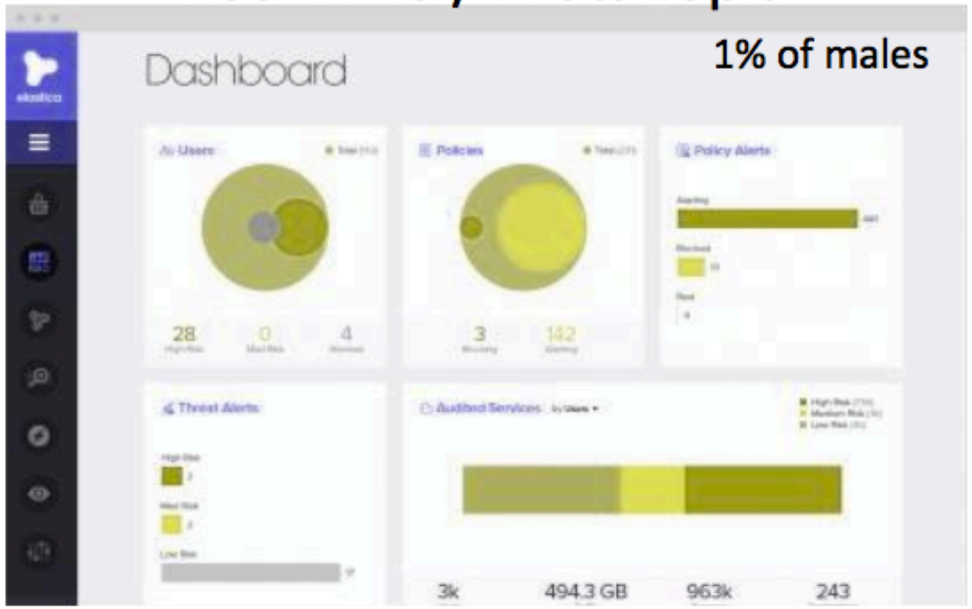
Original

Green-Blind / Deuteranopia

8% of male
0.5% female

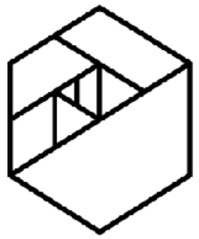
Red-Blind / Protanopia

1% of males



Blue-Blind / Tritanopia

rare

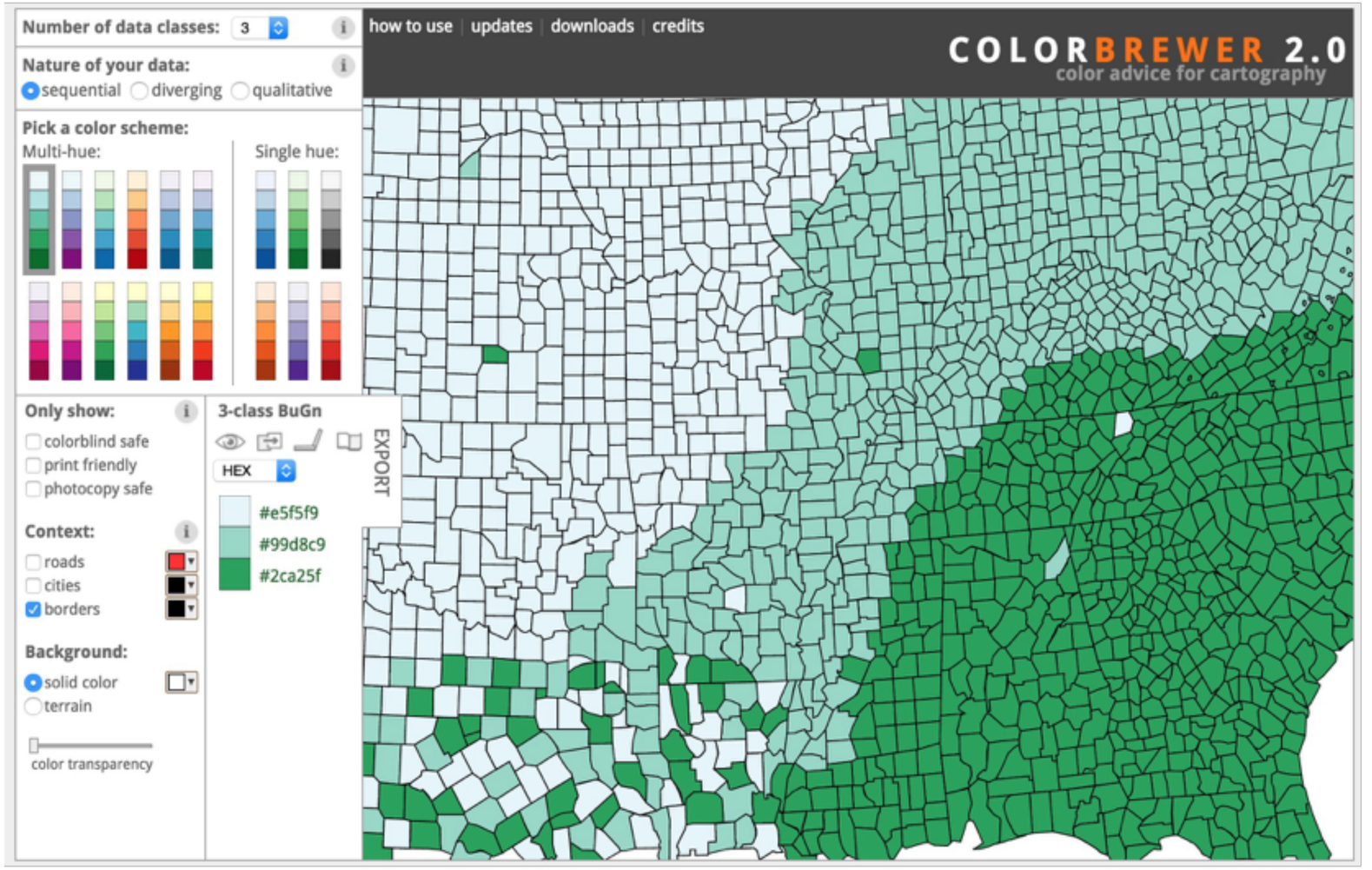


METIS

http://www.color-blindness.com/coblis-color-blindness-simulator/

METIS

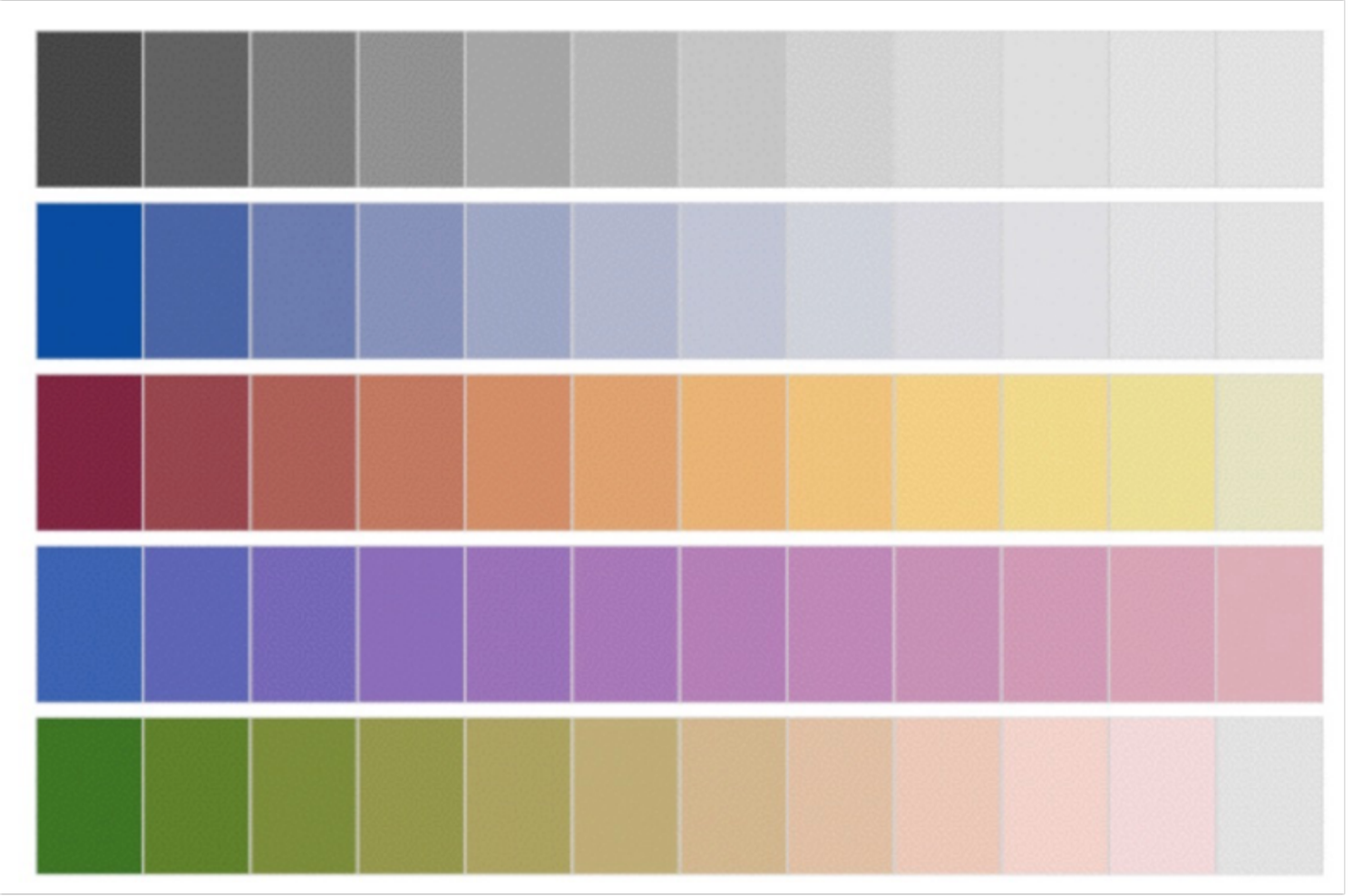
Color



colorbrewer2.org

METIS

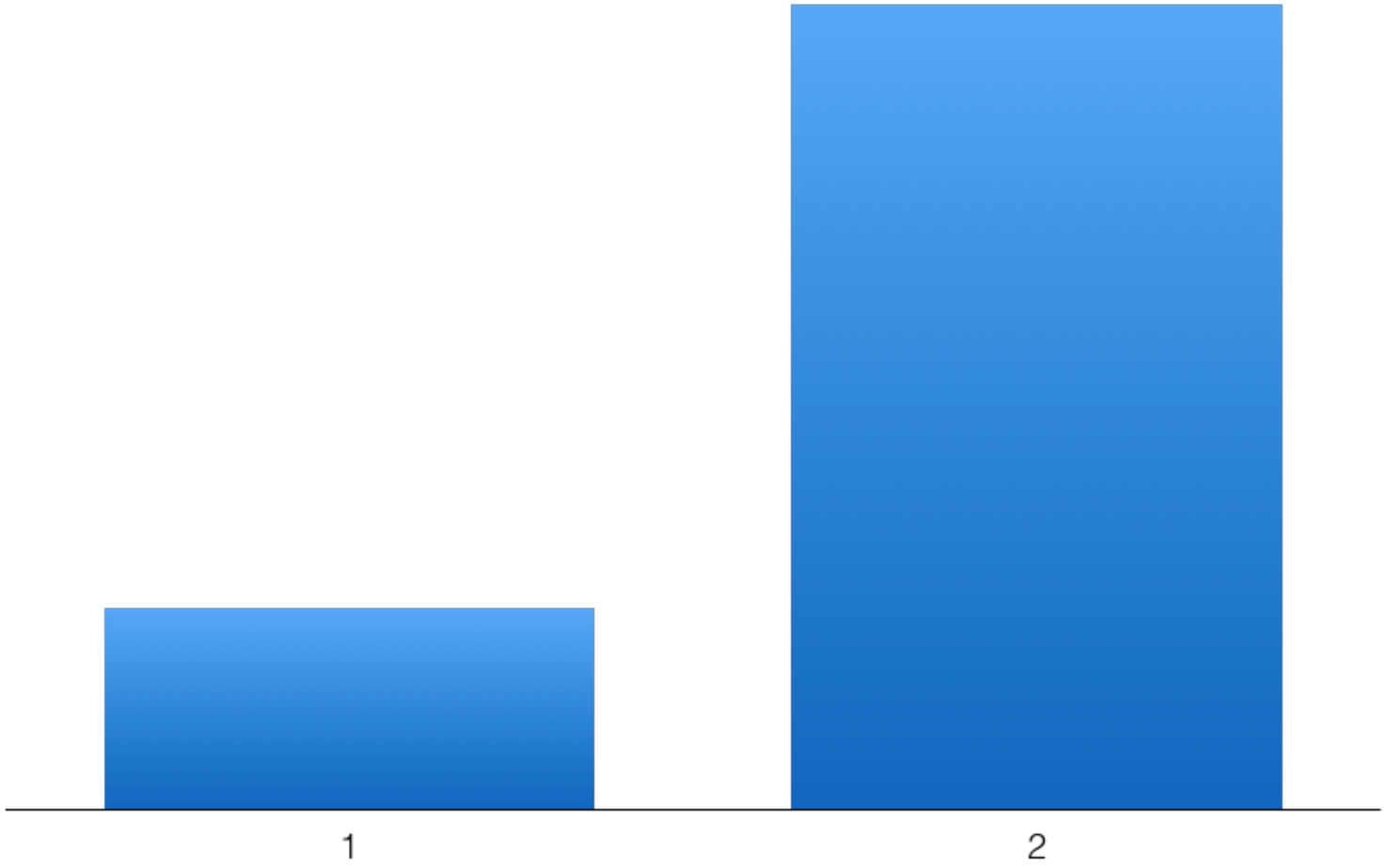
Color

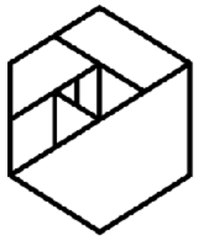


colorbrewer2.org

METIS

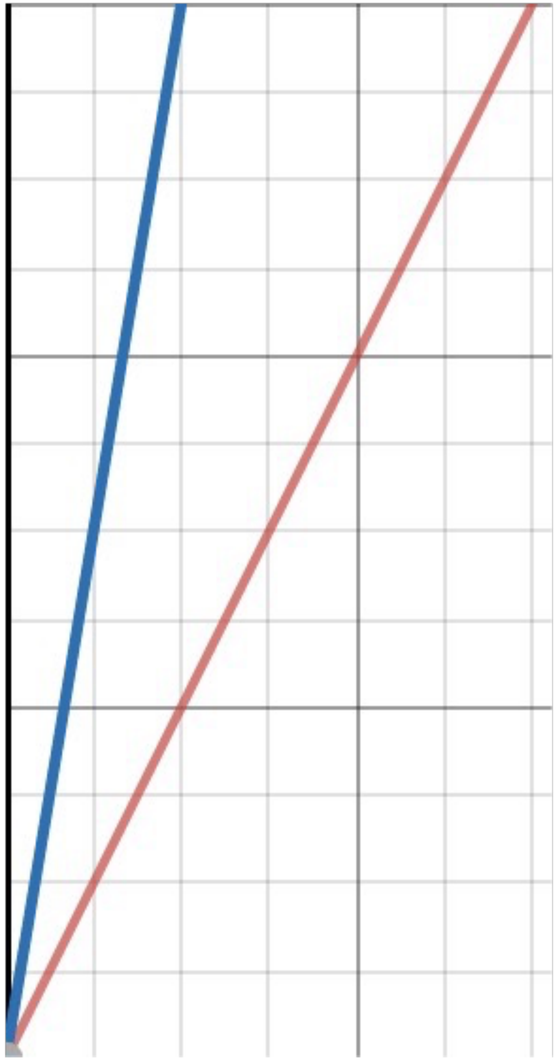
Height Difference?

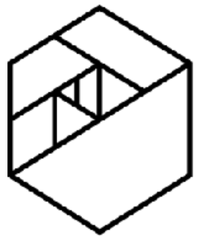




METIS

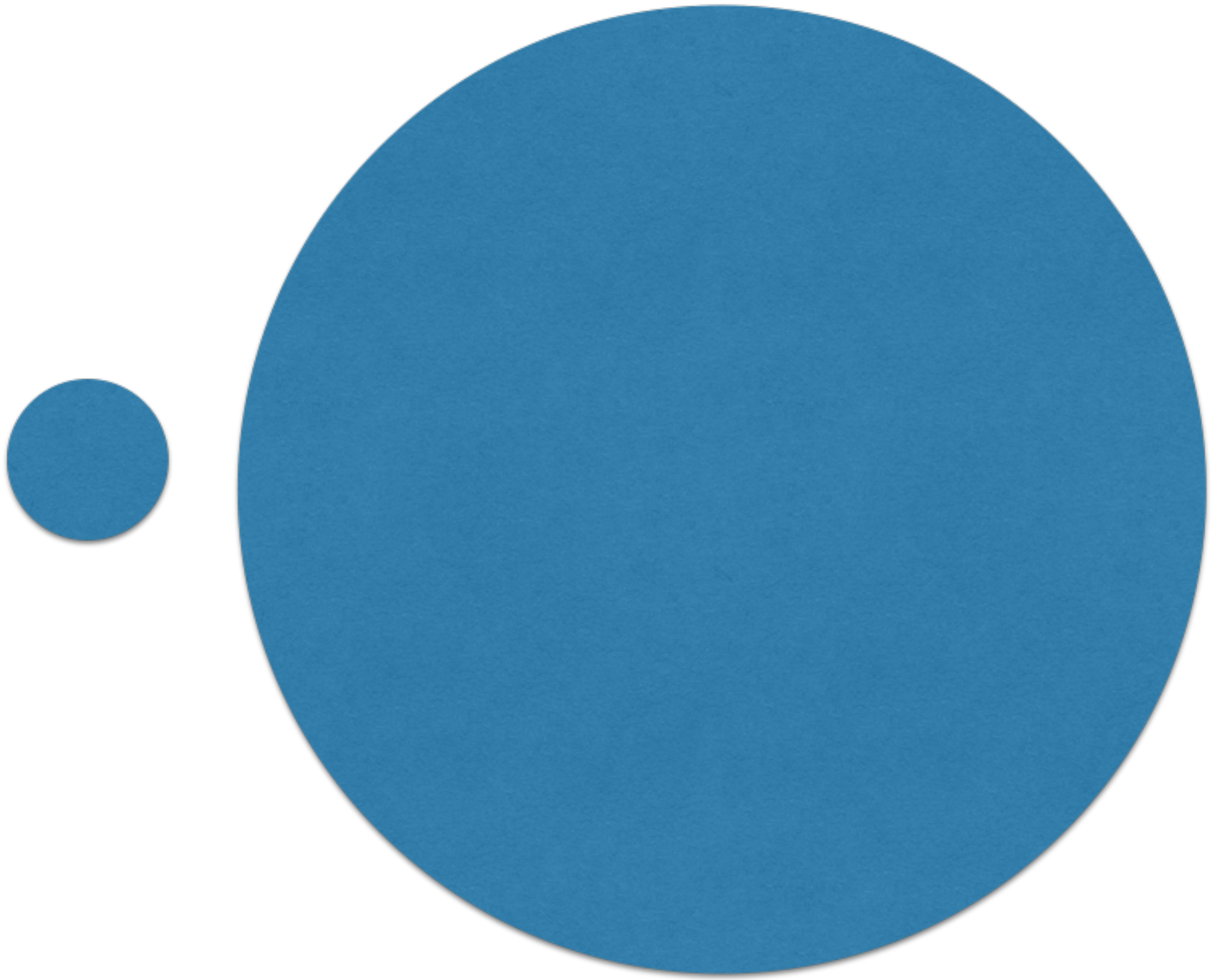
Slope Difference?





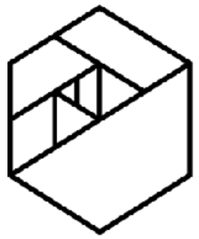
METIS

Area Difference?



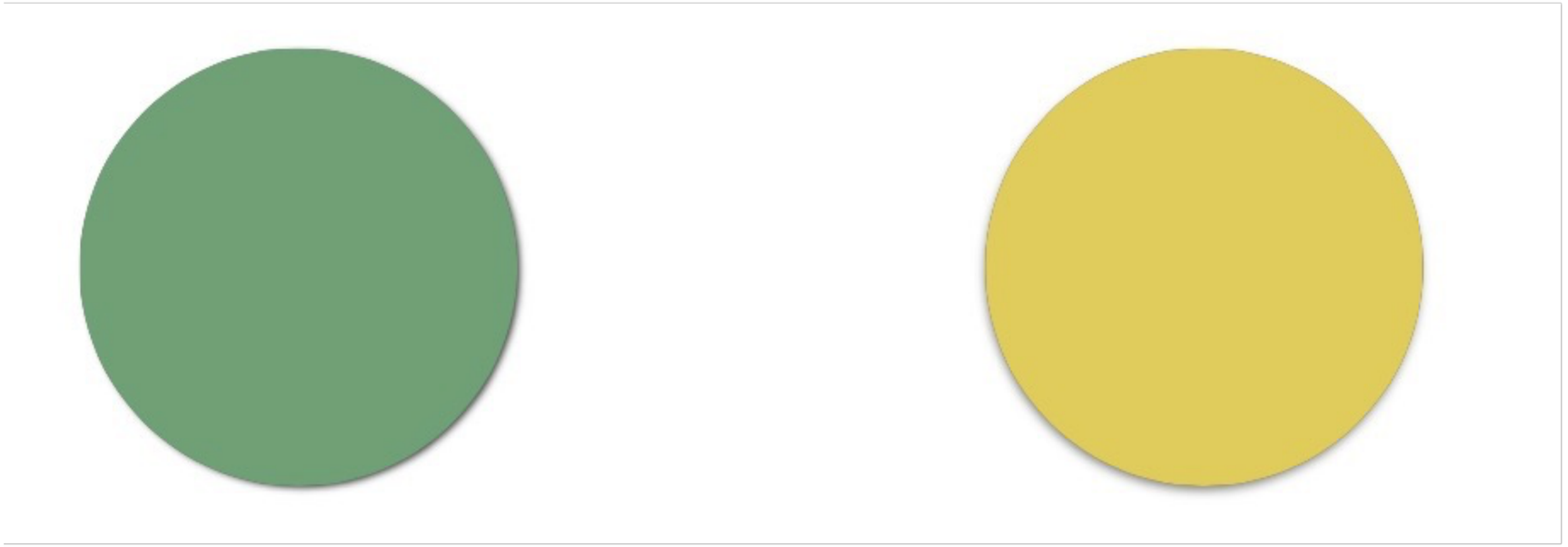
METIS

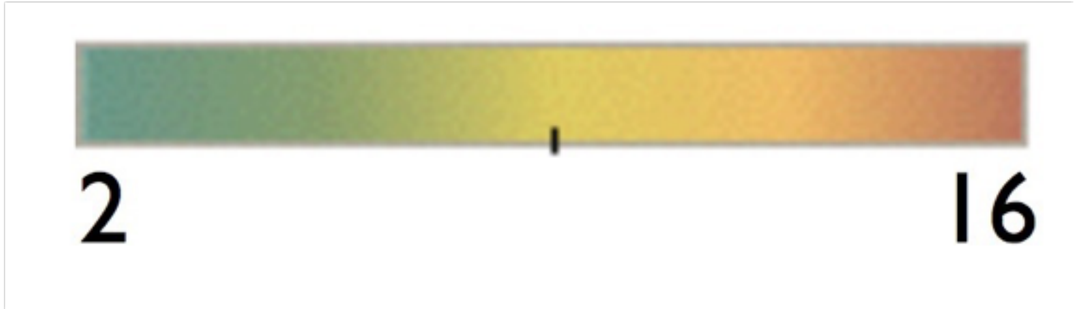
Saturation Difference?



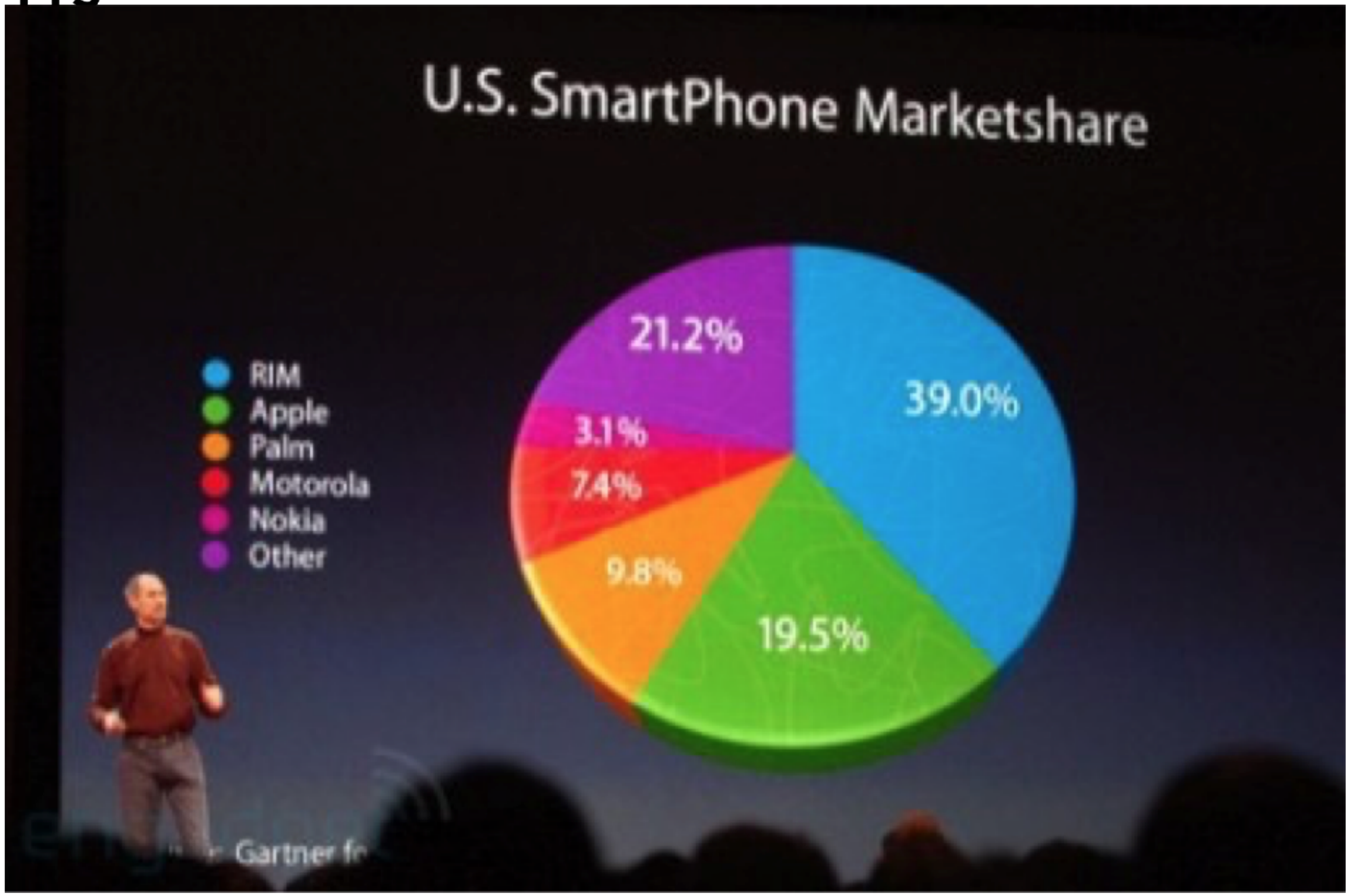
METIS

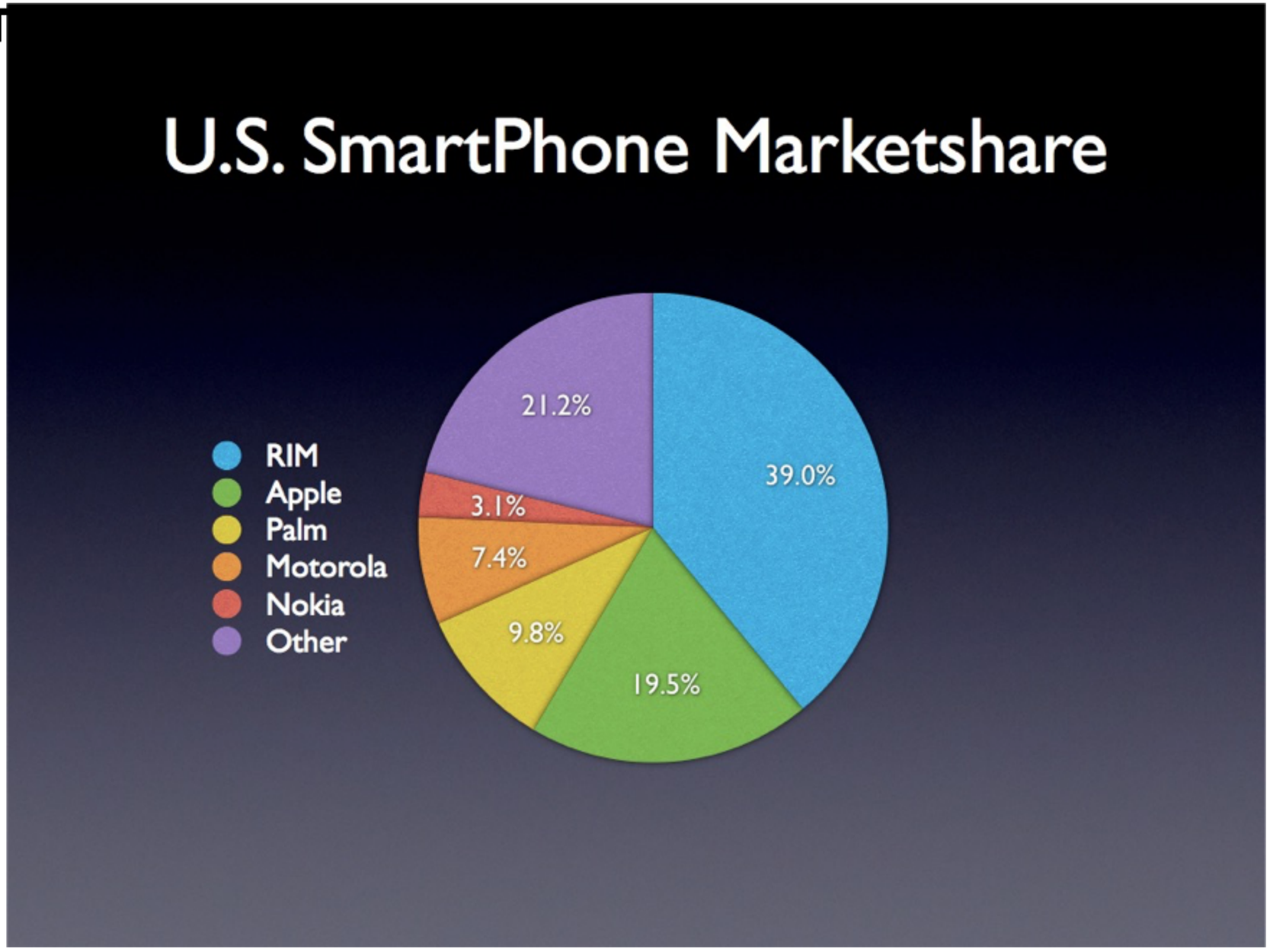
Mapping Difference?





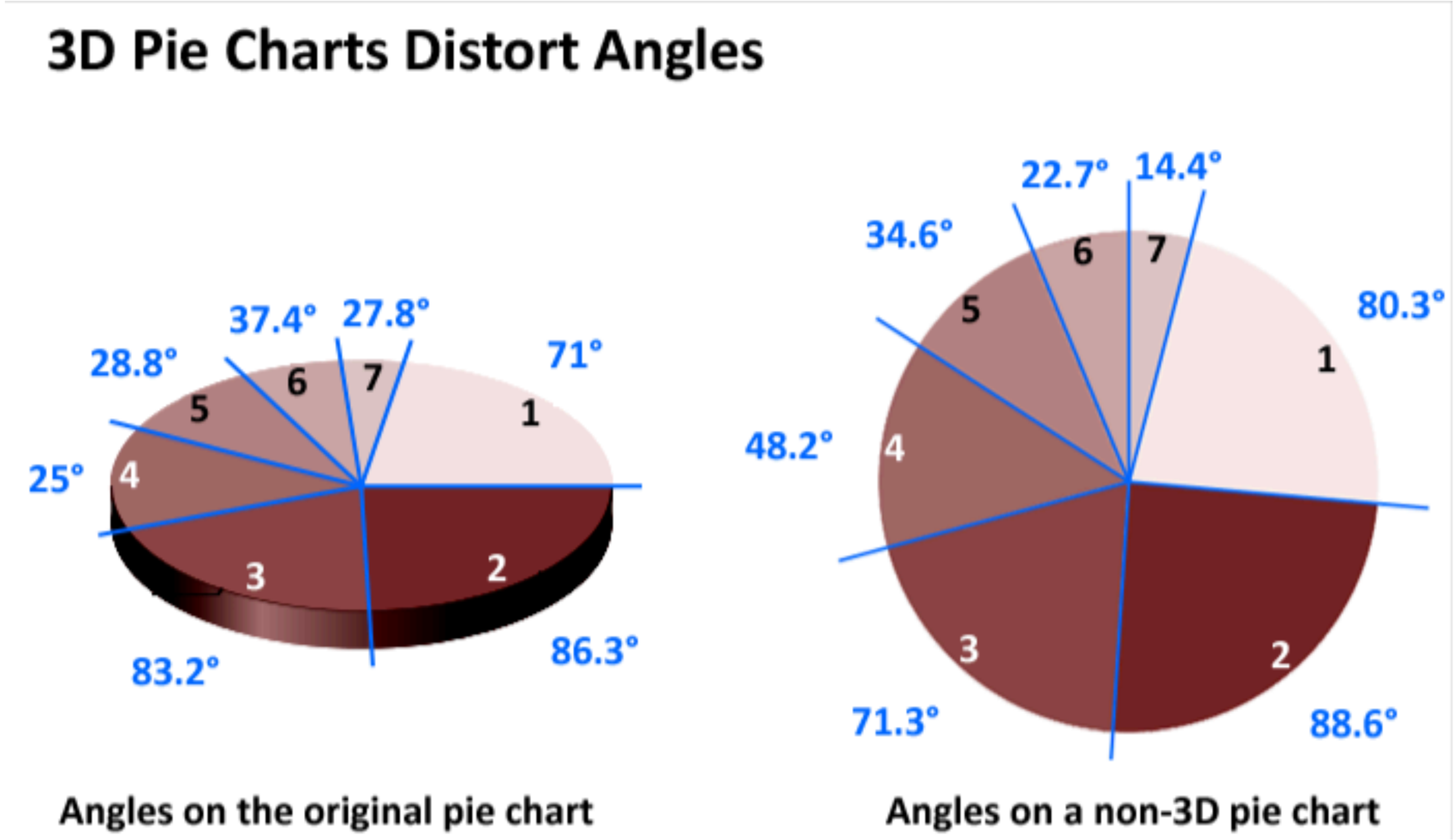
METIS


U.S. SmartPhone Marketshare
RIM
Apple
Palm
Motorola
Nokia
Other
39.0%
19.5%
9.8%
7.4%
3.1%
21.2%
Gartner fr


U.S. SmartPhone Marketshare
RIM
Apple
Palm
Motorola
Nokia
Other
39.0%
19.5%
9.8%
7.4%
3.1%
21.2%

METIS

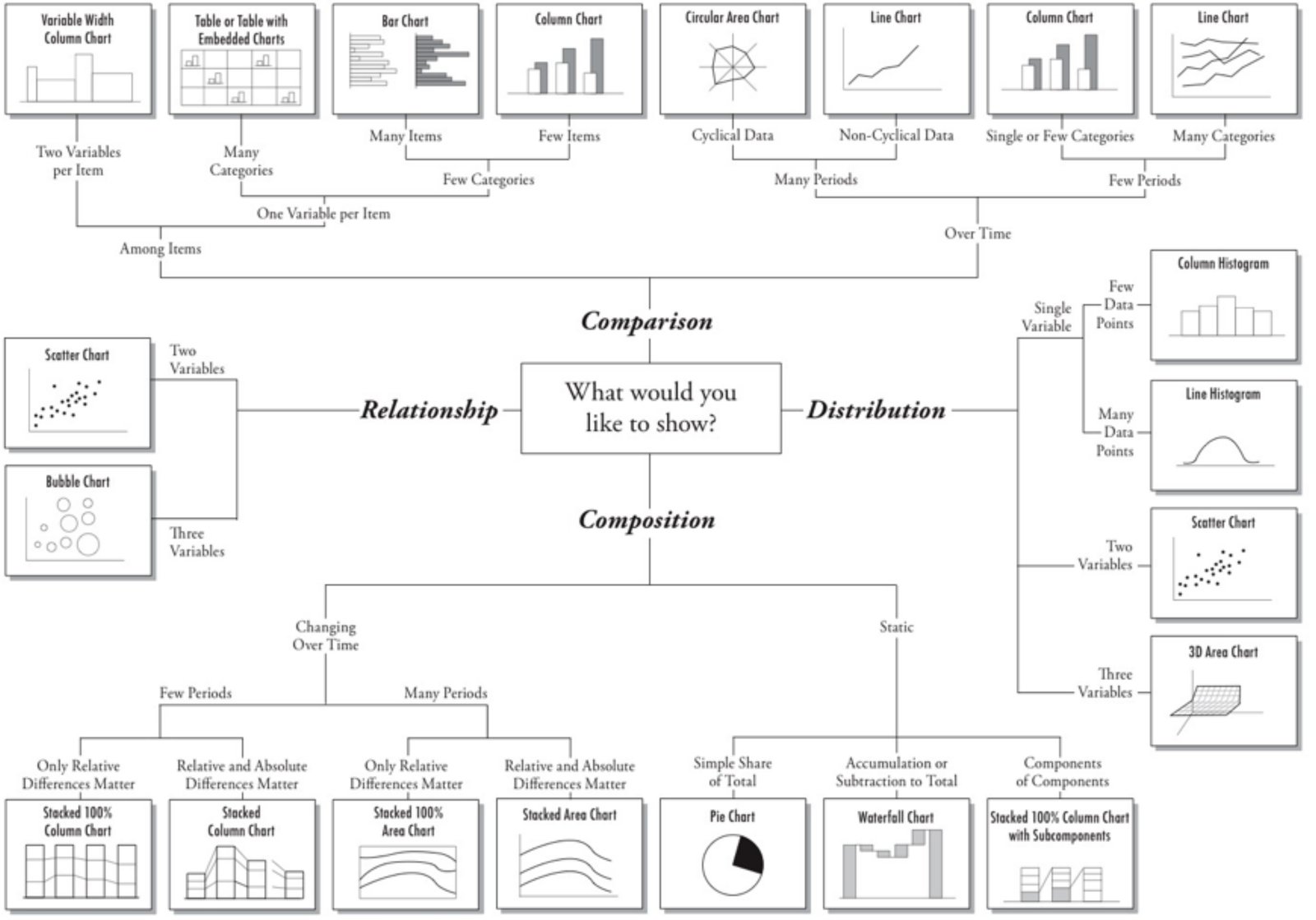
3D Pie Charts are BAD!

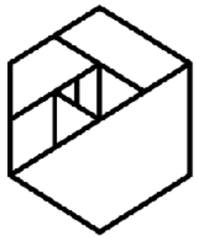


h8p://themendozaline.org/post/113828947026/3d-pie-charts-are-lie-charts

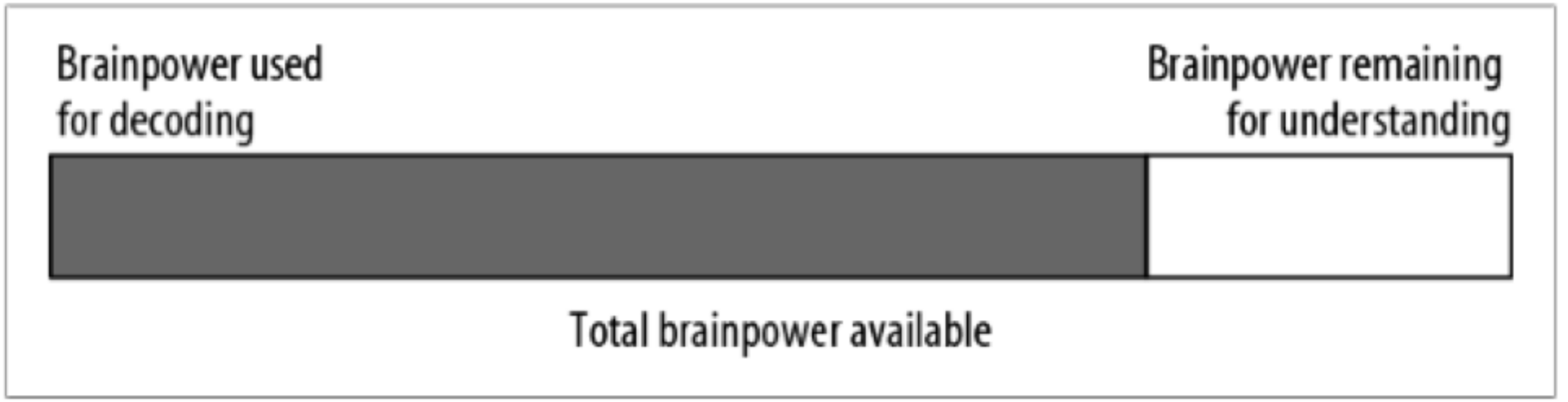
MET

Chart Suggestions—A Thought-Starter

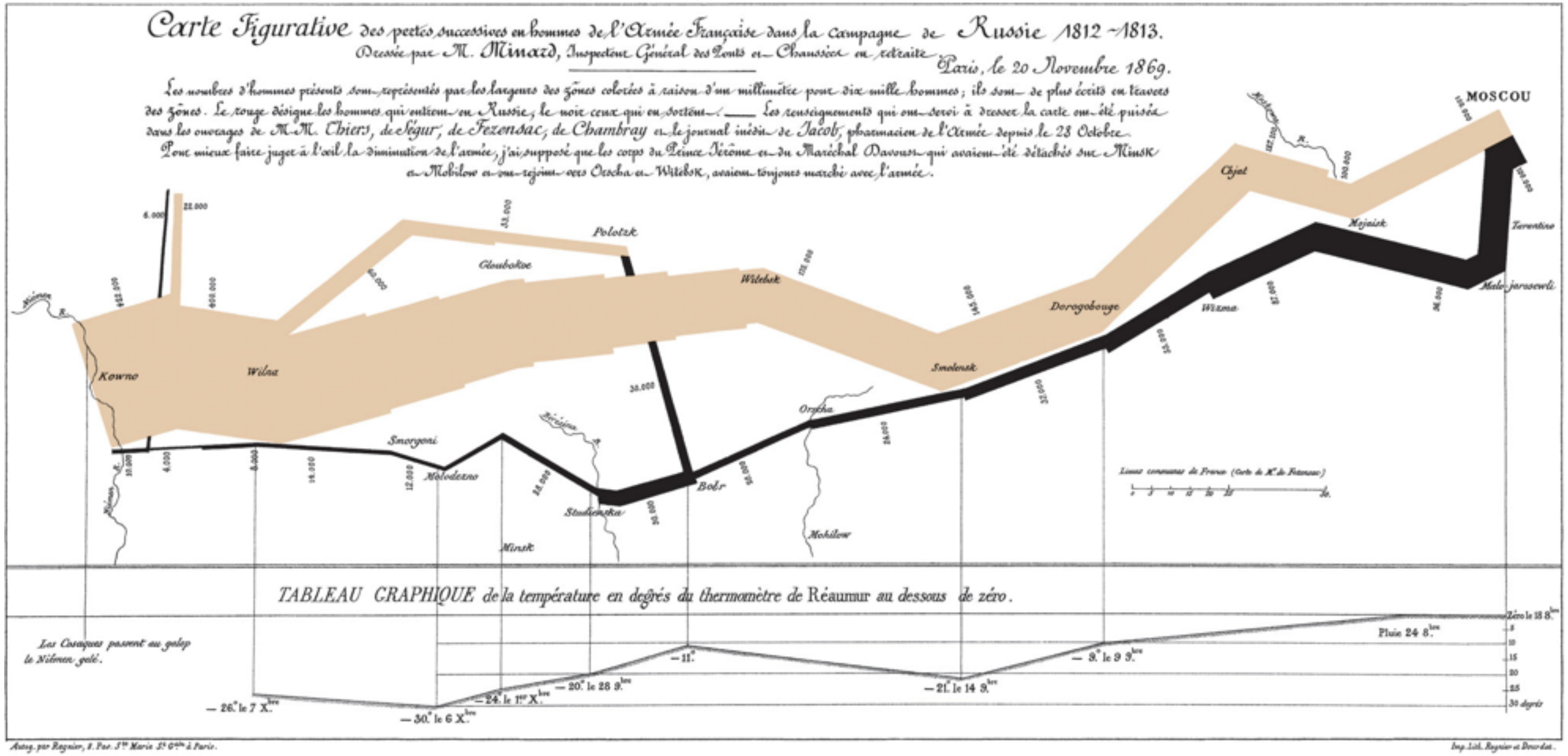




METIS

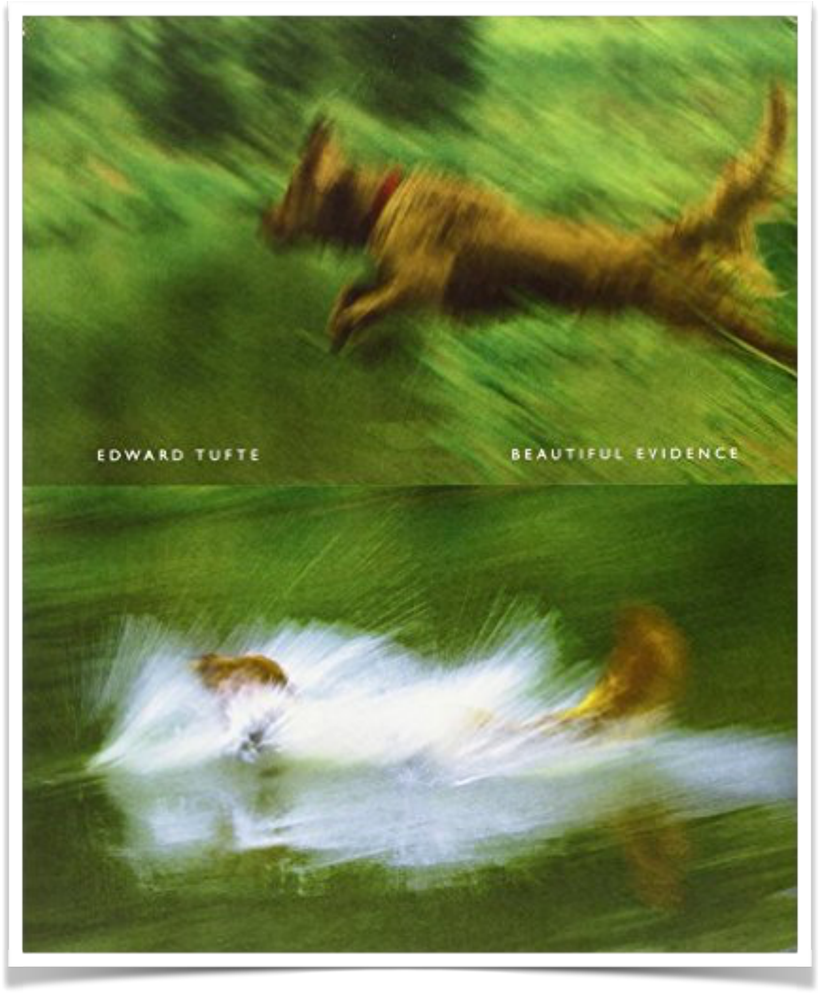

Brainpower used
for decoding
Brainpower remaining
for understanding
Total brainpower available

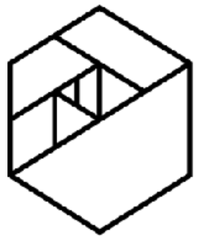
METIS


Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.
Paris, le 20 Novembre 1869.
Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.
Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.
MOSCOU
Kowno
Wilna
Polotzk
Gloubokoe
Witebsk
Smolensk
Dorogobouge
Chjat
Mojaisk
Tarantino
Malo-jarosewli
Wizma
Orscha
Bobr
Studienska
Minsk
Mohilow
Smorgoni
Molodezno
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.
Les Cosaques passent au galop le Niémen gelé.
Zéro le 18 8.bre
Pluie 24 8.bre
— 9° le 9 9.bre
— 21° le 14 9.bre
— 11°
— 20° le 28 9.bre
— 24° le 1er X.bre
— 30° le 6 X.bre
— 26° le 7 X.bre
30 degrés

METIS

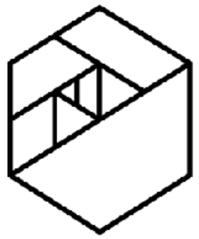
Recommended Reading





METIS

Data Visualization in Python



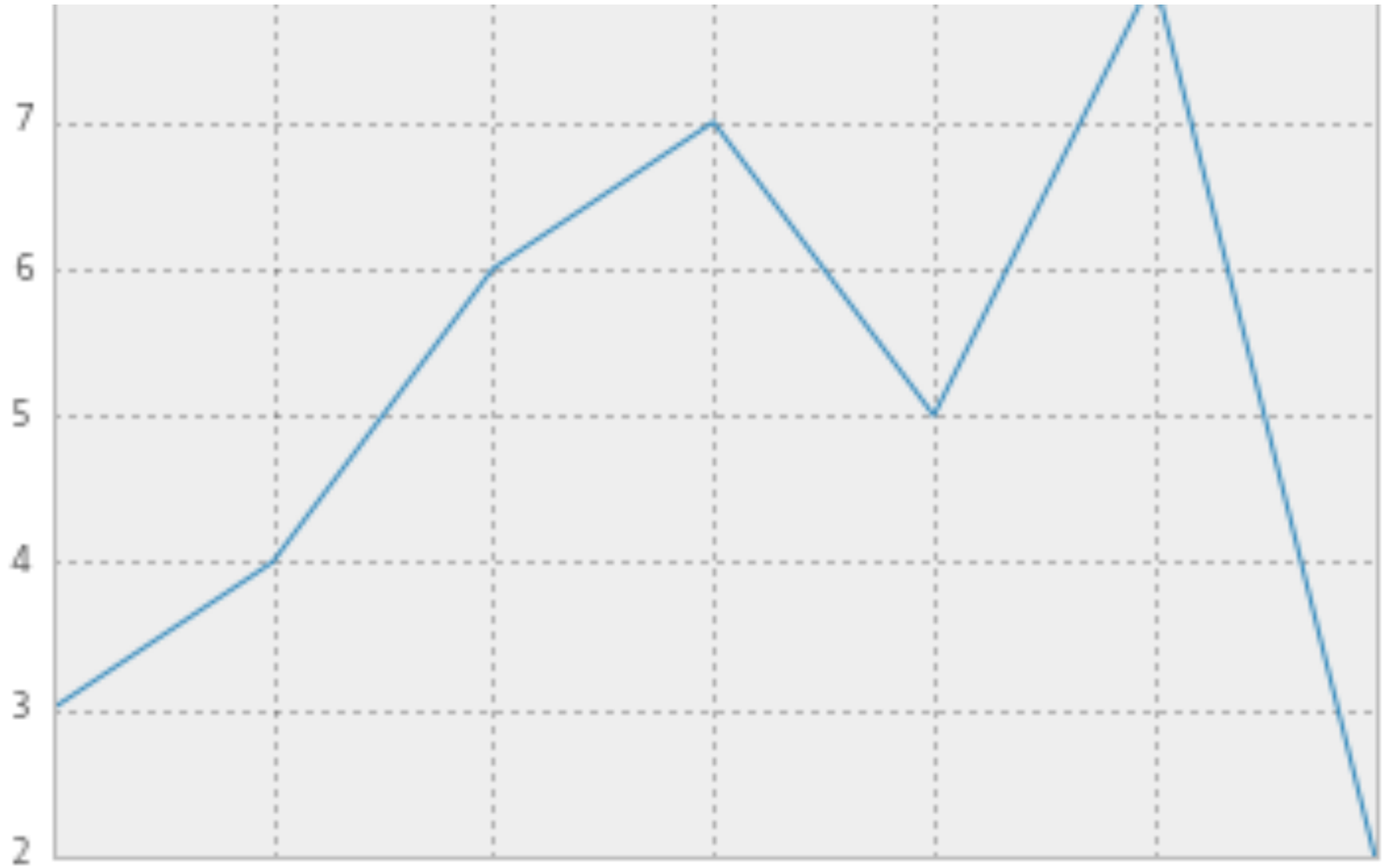
METIS

Python Visualization Libraries

- **Matplotlib**: “Lowest” level visualization library. Very powerful, but little abstraction
- **Pandas/Pyplot**: Easier, but requires transformation for complex visualizations
- **Seaborn:** Good for advanced statistical visualizations
- **Altair:** New library: declarative statistical visualization library

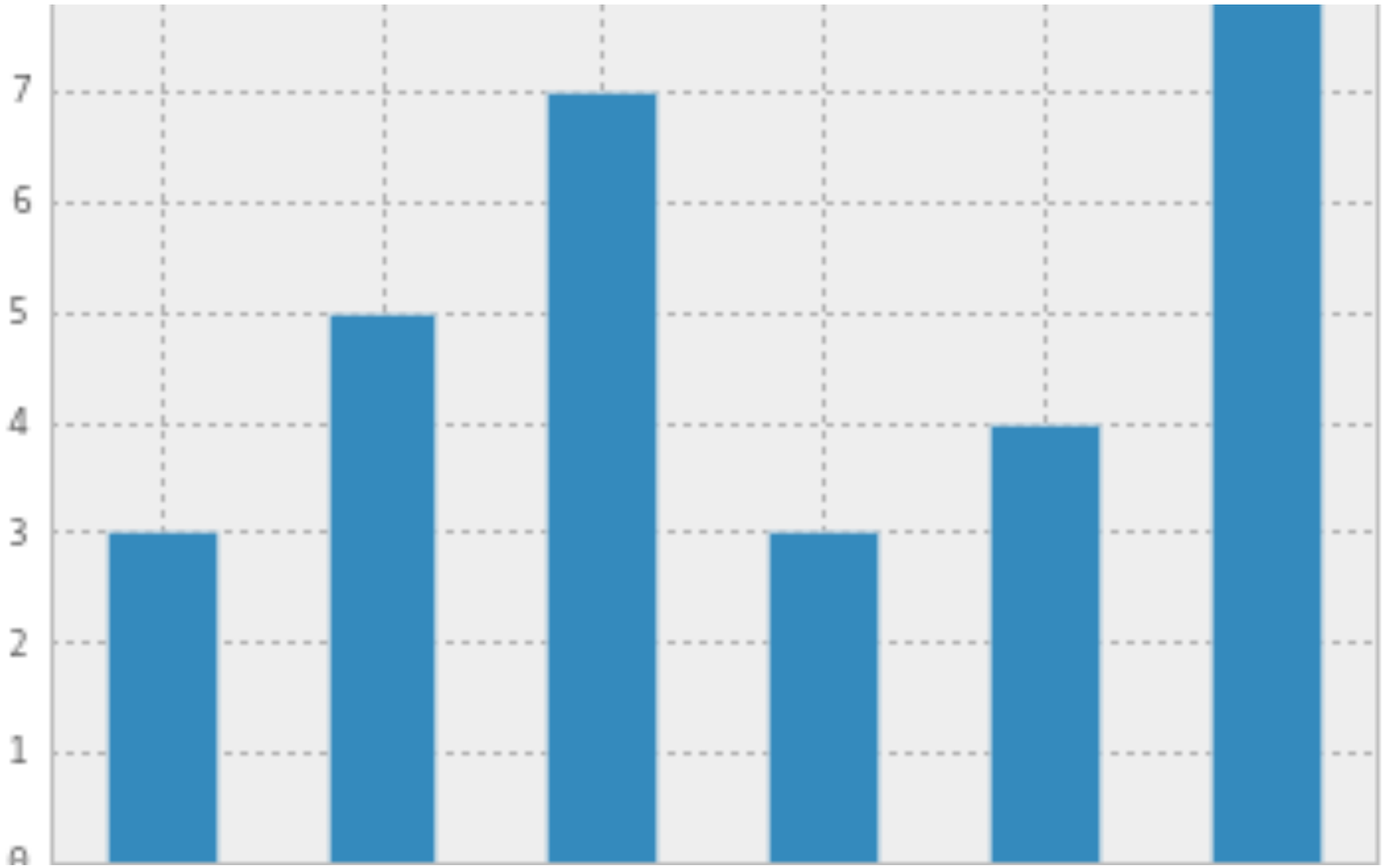
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline
pd.options.display.mpl_style = 'default'
```

METIS



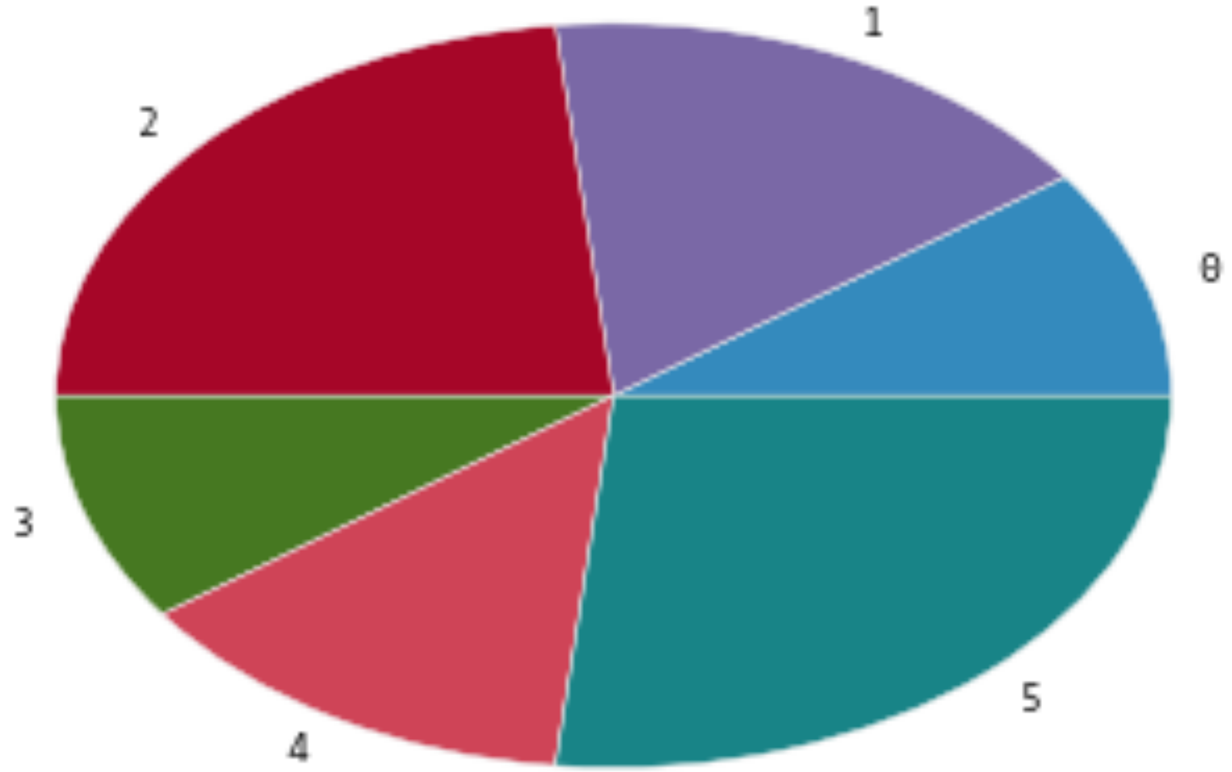

```
data = pd.Series( [3,4,6,7,5,8,2 ])
graph = data.plot()
```

METIS



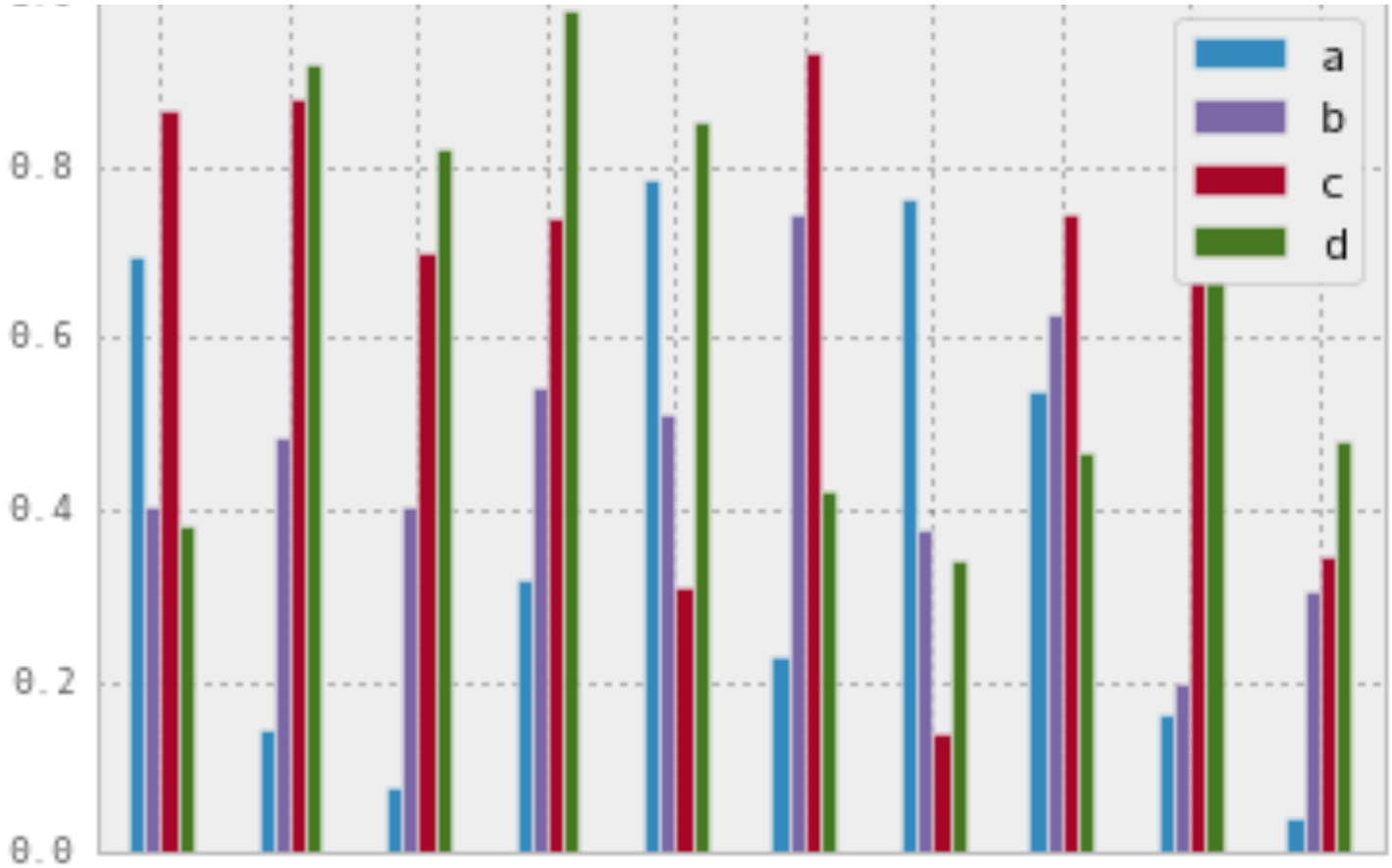

```
barchart = data.plot( kind="bar" )
```

METIS



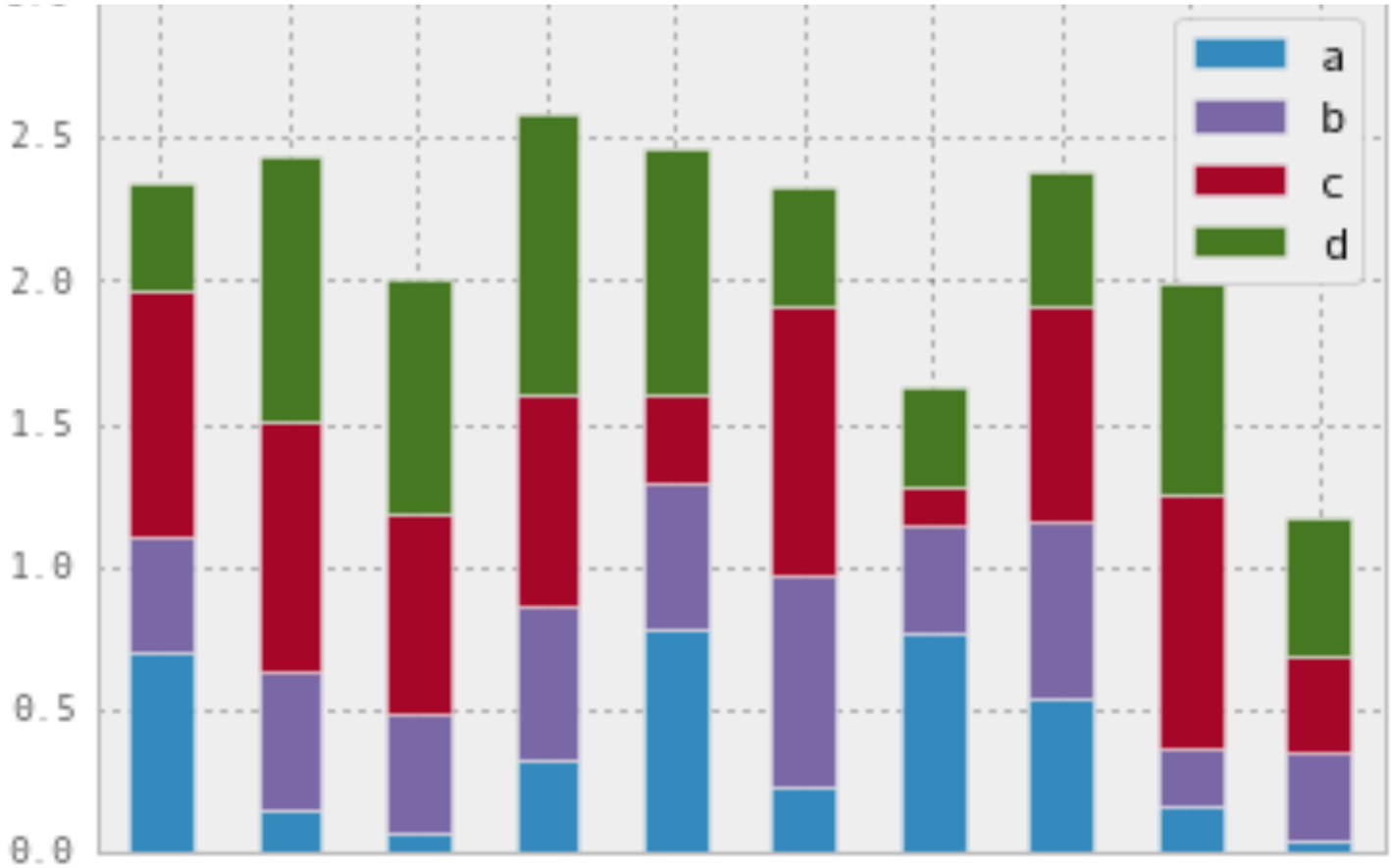

```
piechart = data.plot( kind="pie" )
```

METIS

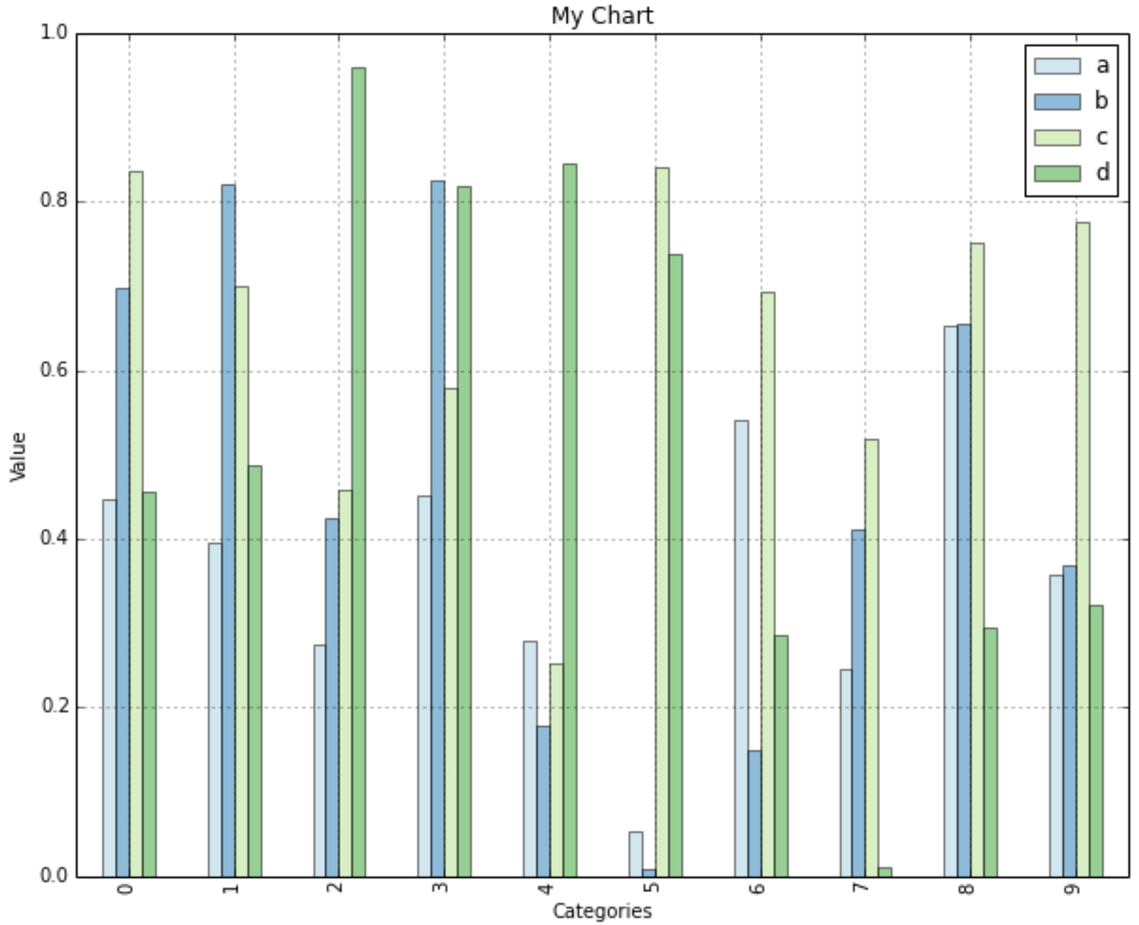



```
df2 = pd.DataFrame(np.random.rand(10, 4),
columns=['a', 'b', 'c', 'd'])
df2.plot( kind='bar' )
```

METIS

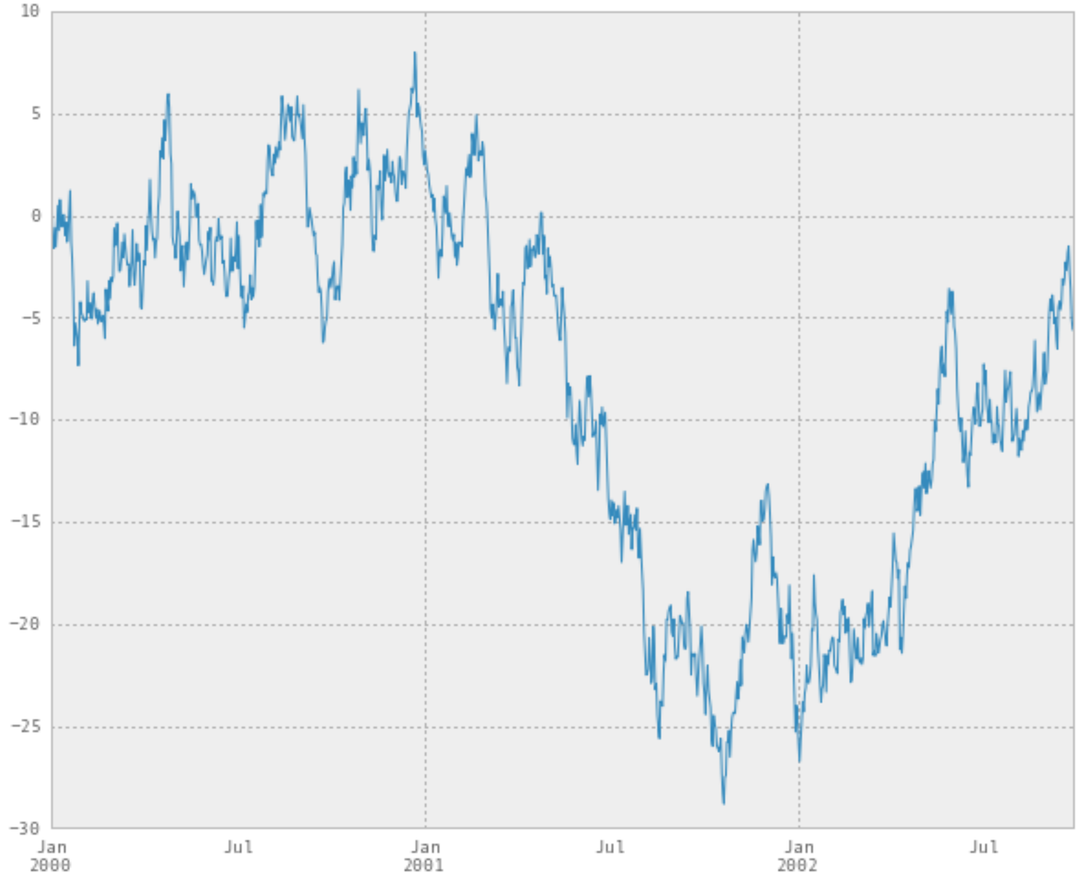



```
df2.plot( kind='bar', stacked=True )
```



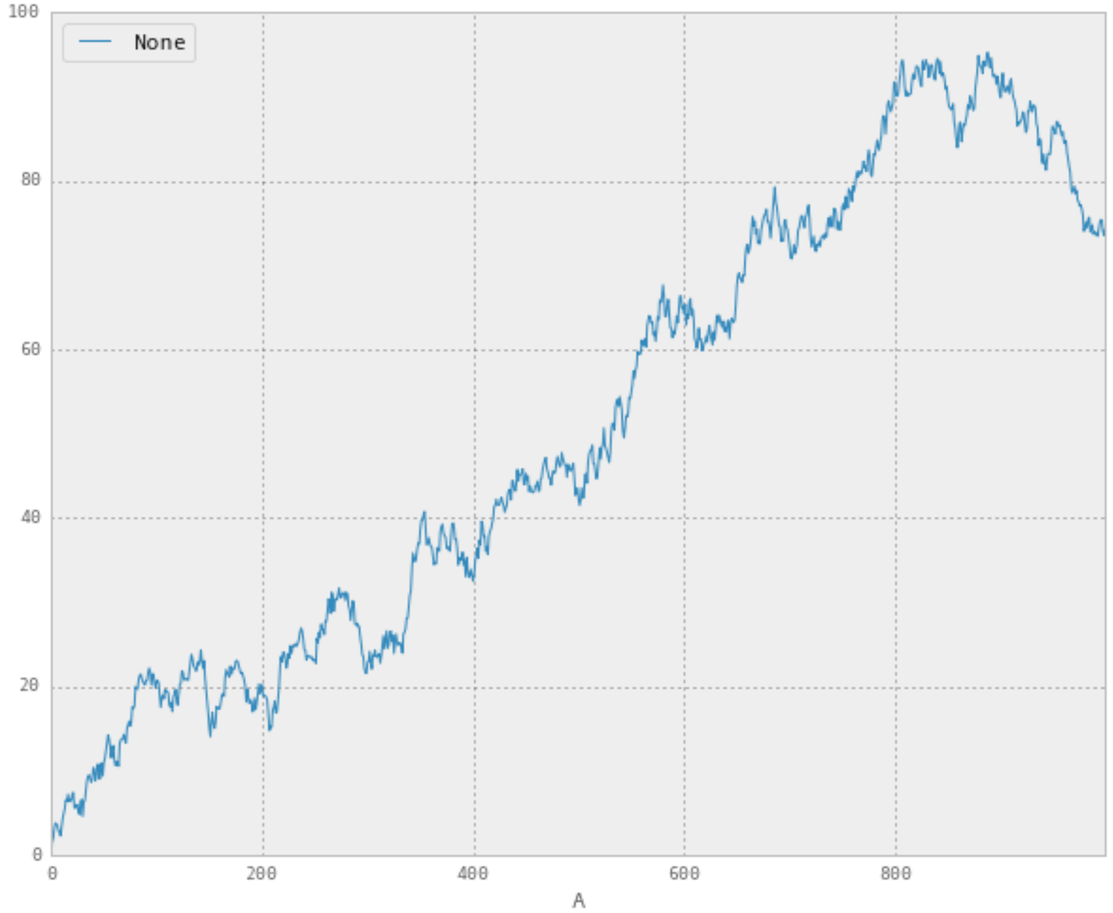

```
df2.plot( kind='bar',
          color=('#a6cee3','#1f78b4','#b2df8a','#33a02c' ),
           alpha=0.5,
             width=0.5,
             figsize=(10, 8))
plt.title( "My Chart" )
plt.xlabel( "Categories" )
plt.ylabel( "Value" )
```

METIS

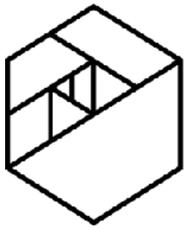



```
ts = pd.Series(np.random.randn( 1000),
index=pd.date_range('1/1/2000', periods=1000))
ts = ts.cumsum()
timeseriesChart = ts.plot( figsize=(10, 8) )
```

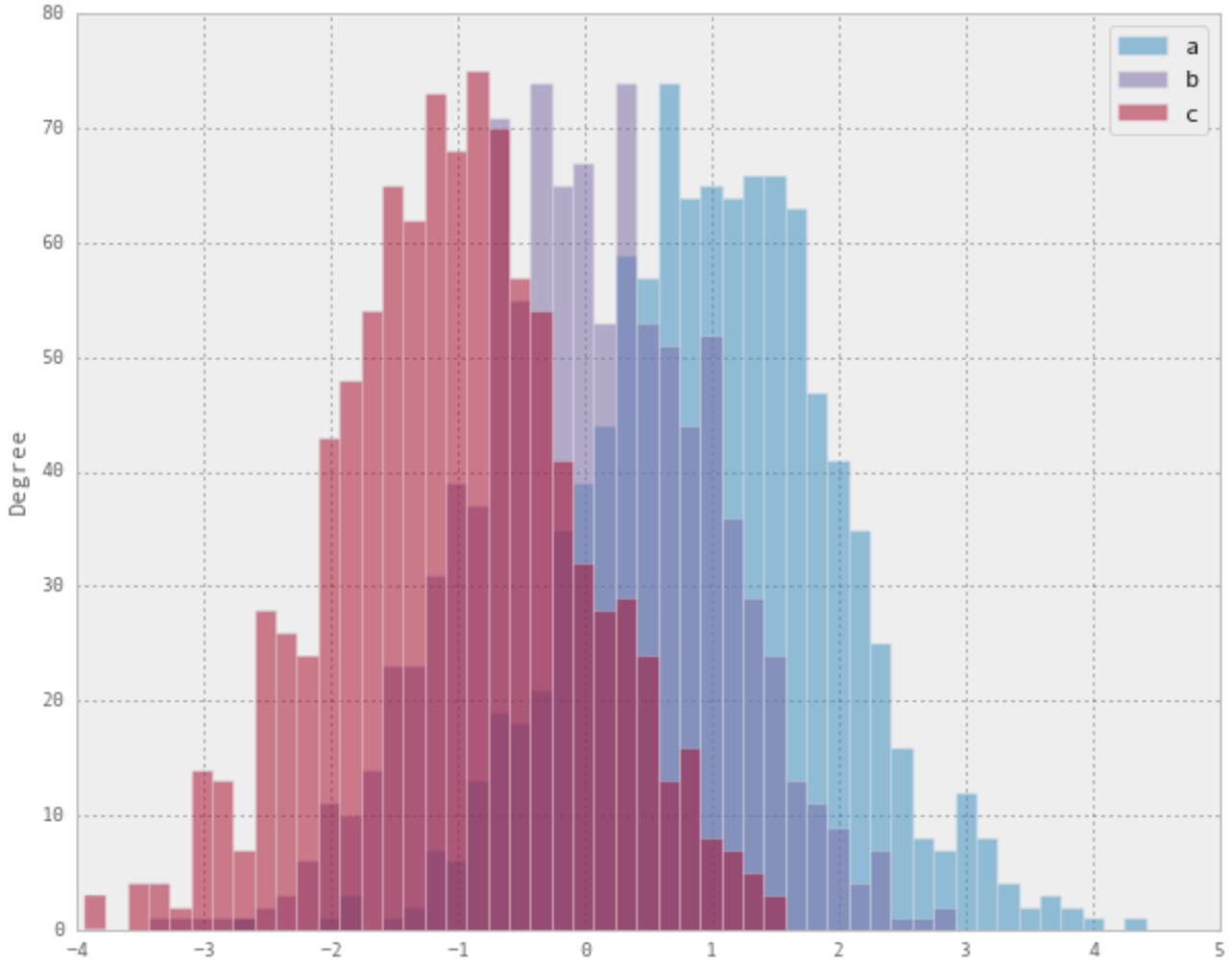
METIS




```
df3 = pd.DataFrame(np.random.randn(1000, 2),
columns=['B', 'C']).cumsum()
df3['A'] = pd.Series(list(range(len(df3))))

df3.plot( x='A', y='B' )
```

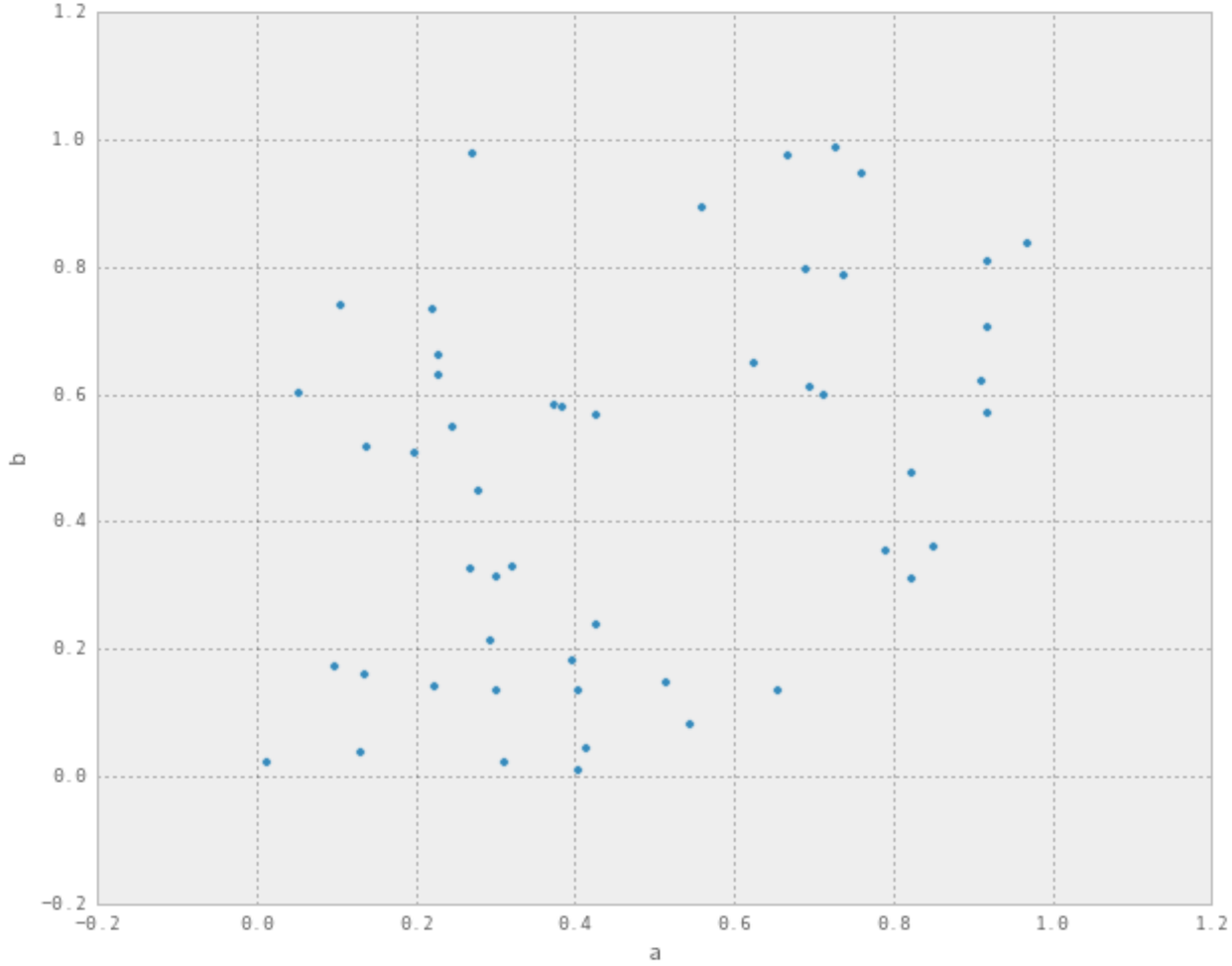


METIS

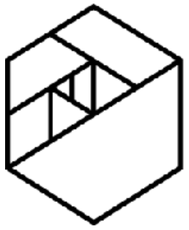



```
df4.plot(kind='hist',
         alpha=0.5,
         bins=50 )
```

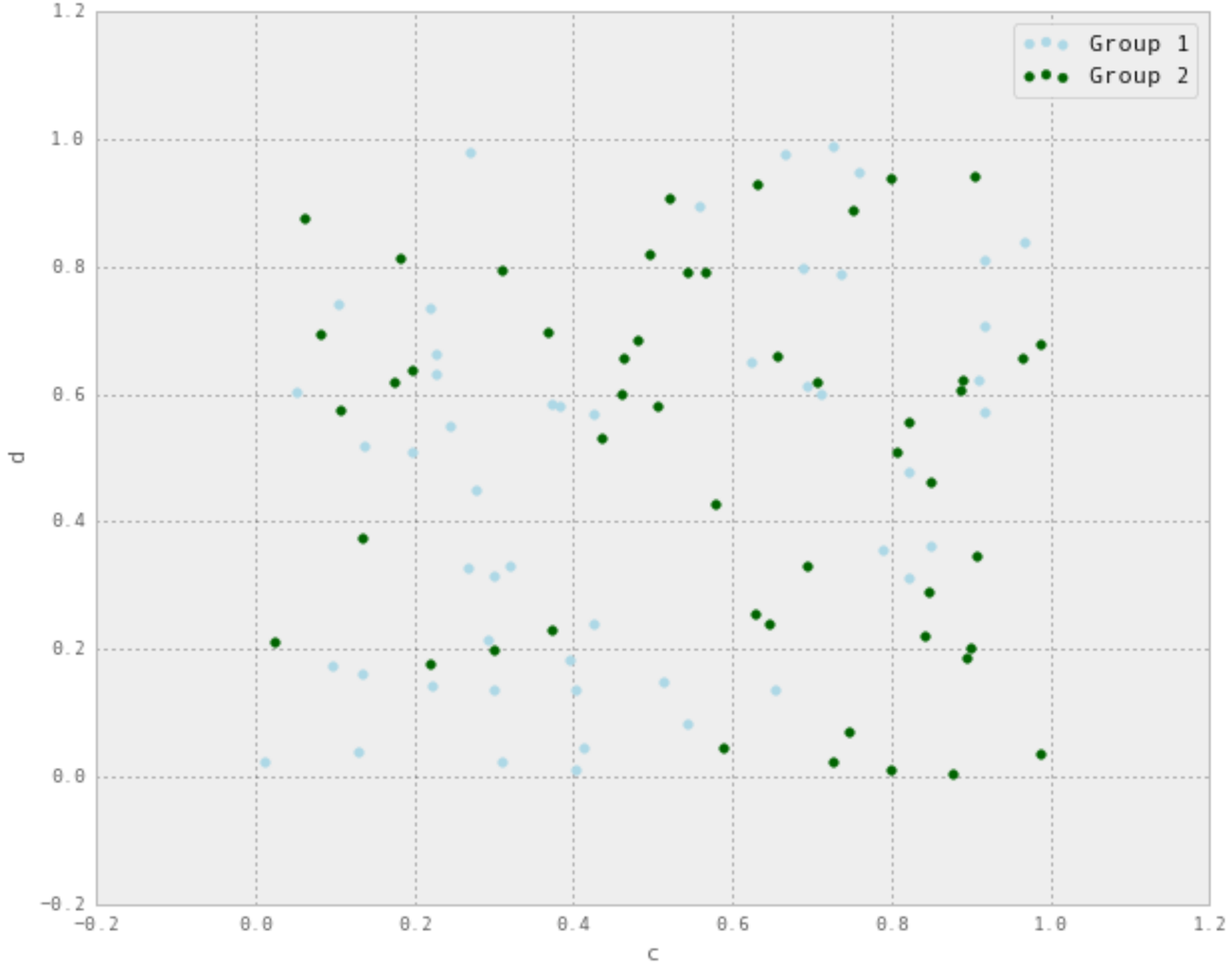
METIS



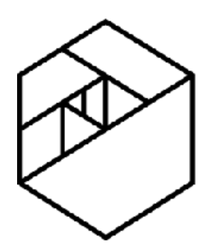

```
df5 = pd.DataFrame(np.random.rand(50, 4),
columns=['a', 'b', 'c', 'd'] )
df5.plot(kind='scatter', x='a', y='b' )
```



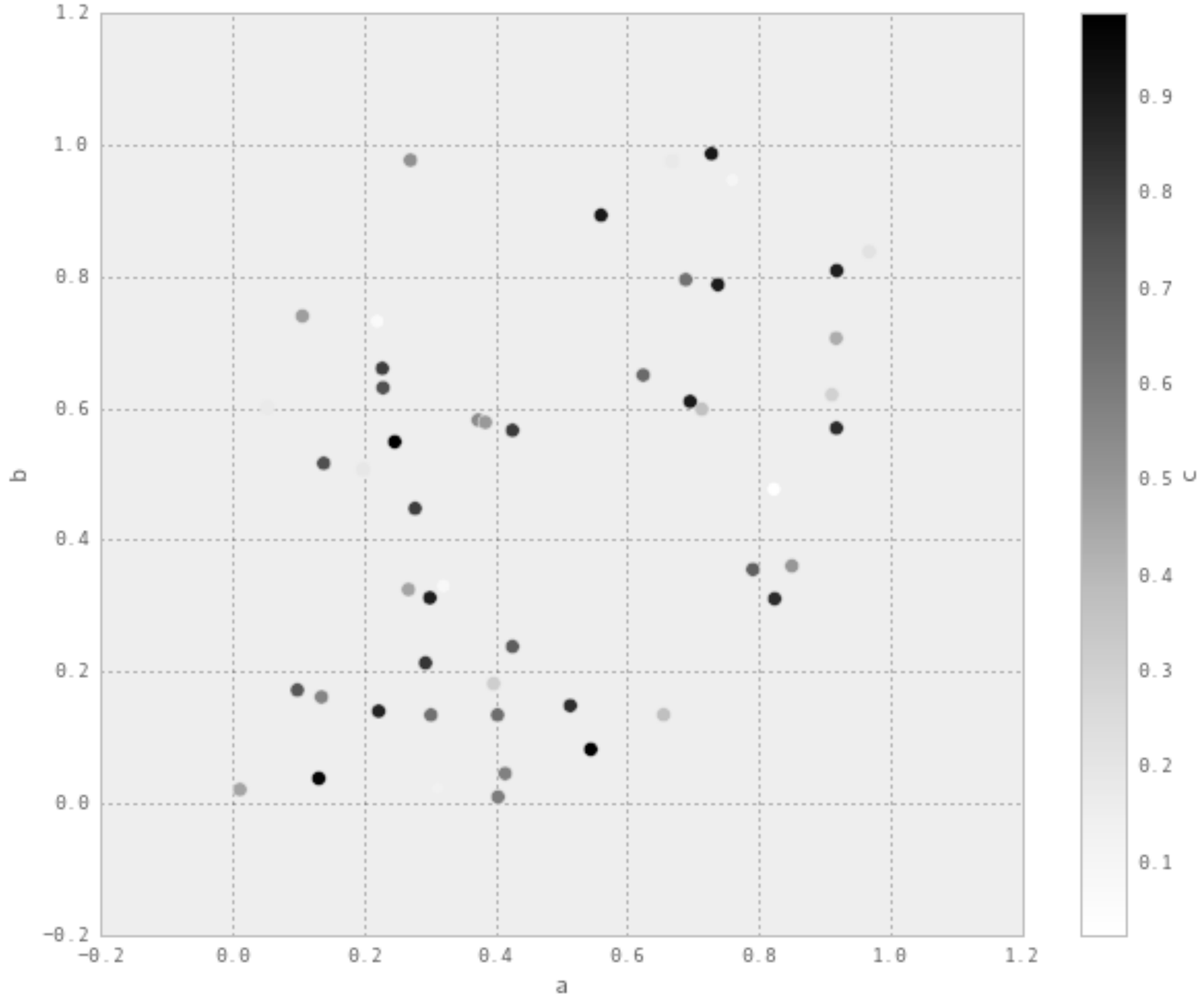
METIS




```
ax = df5.plot(kind='scatter', x='a', y='b',
              color='LightBlue',
              label='Group 1',
              figsize=(10, 8) )

df5.plot(kind='scatter', x='c', y='d',
             color='DarkGreen',
            label='Group 2',
            ax=ax)
```

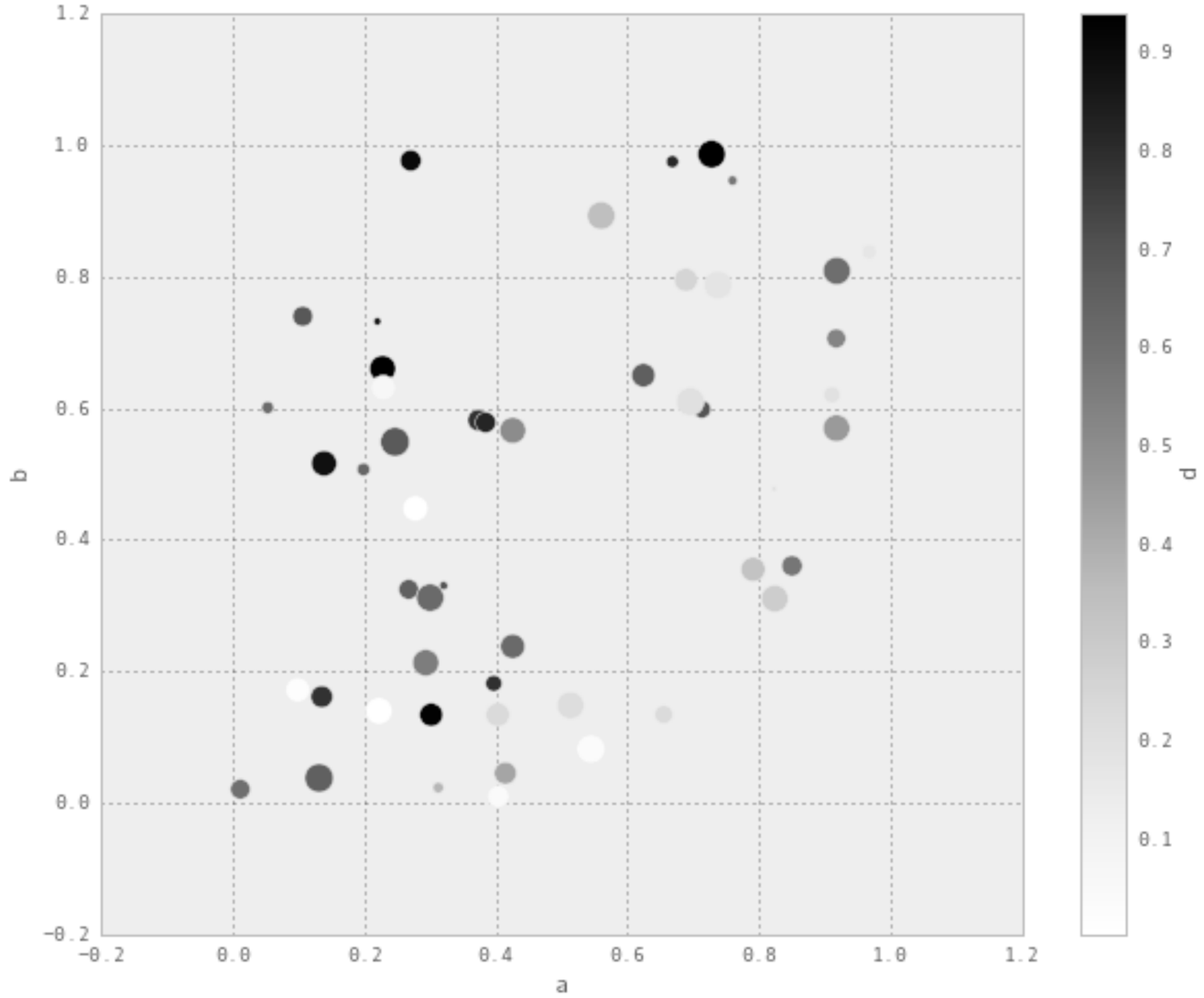


METIS



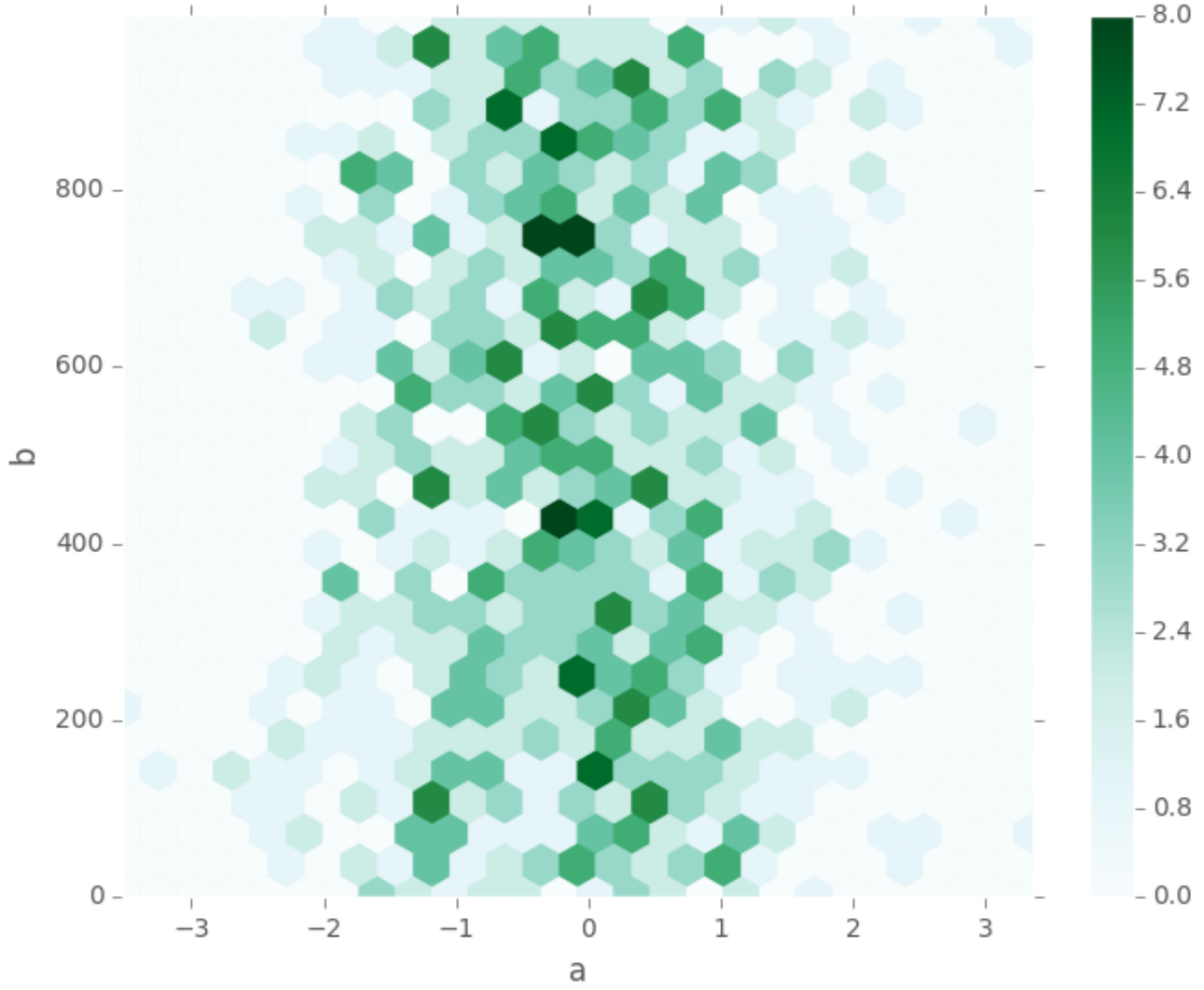

```
df5.plot(kind='scatter', x='a', y='b',
c='c', s=50 )
```

METIS

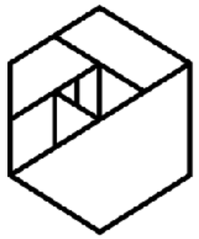



```
scatter = df5.plot(kind='scatter',
                   x='a',
                   y='b',
                   s=df5['c']*200,
                   c='d')
```

METIS

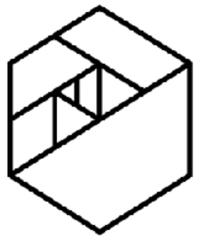



```
df.plot(kind='hexbin', x='a', y='b',
gridsize=25)
```



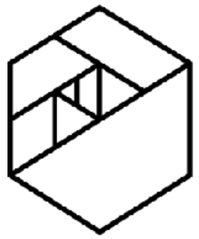
METIS

```
scatterPlot.get_figure().savefig( "scatterPlot.png" )
```



METIS

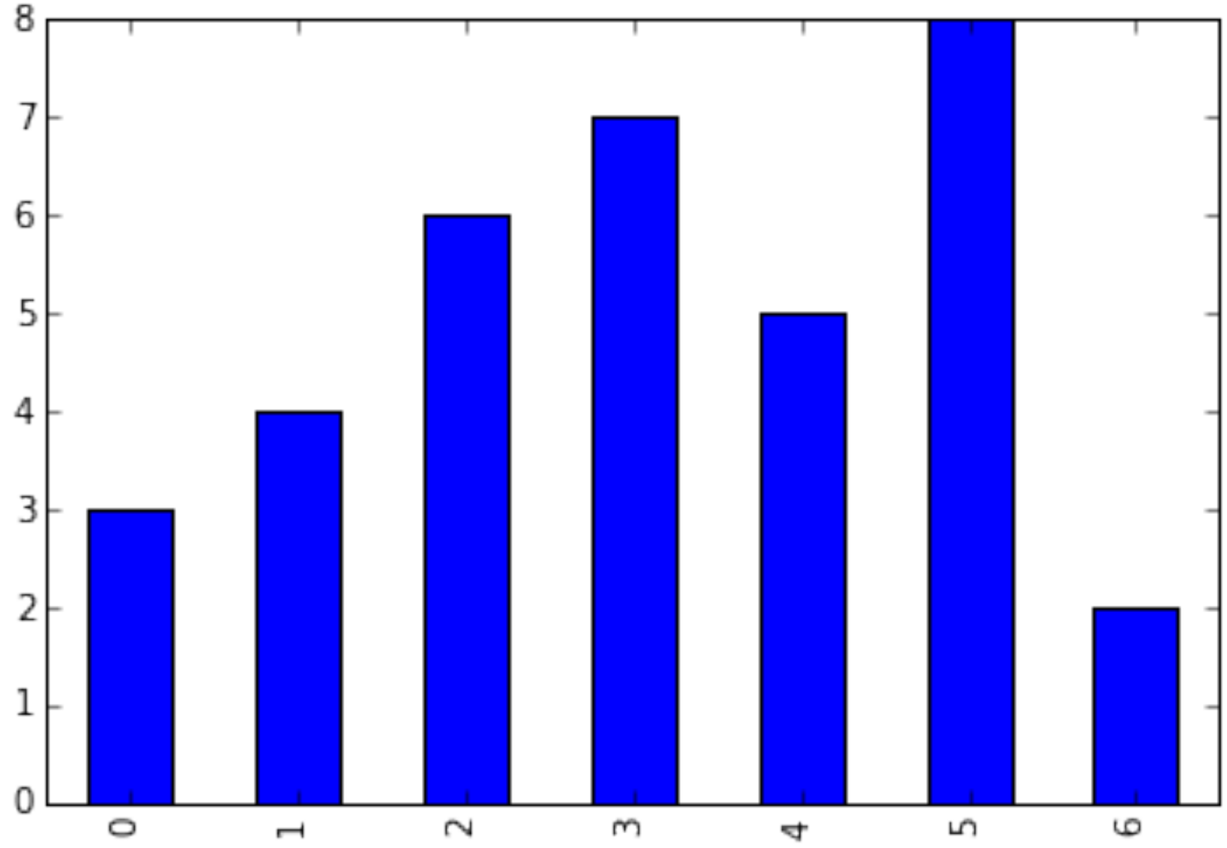
```
print(plt.style.available)

['dark_background', 'grayscale', 'ggplot', 'bmh',
                    'fivethirtyeight']
```



METIS

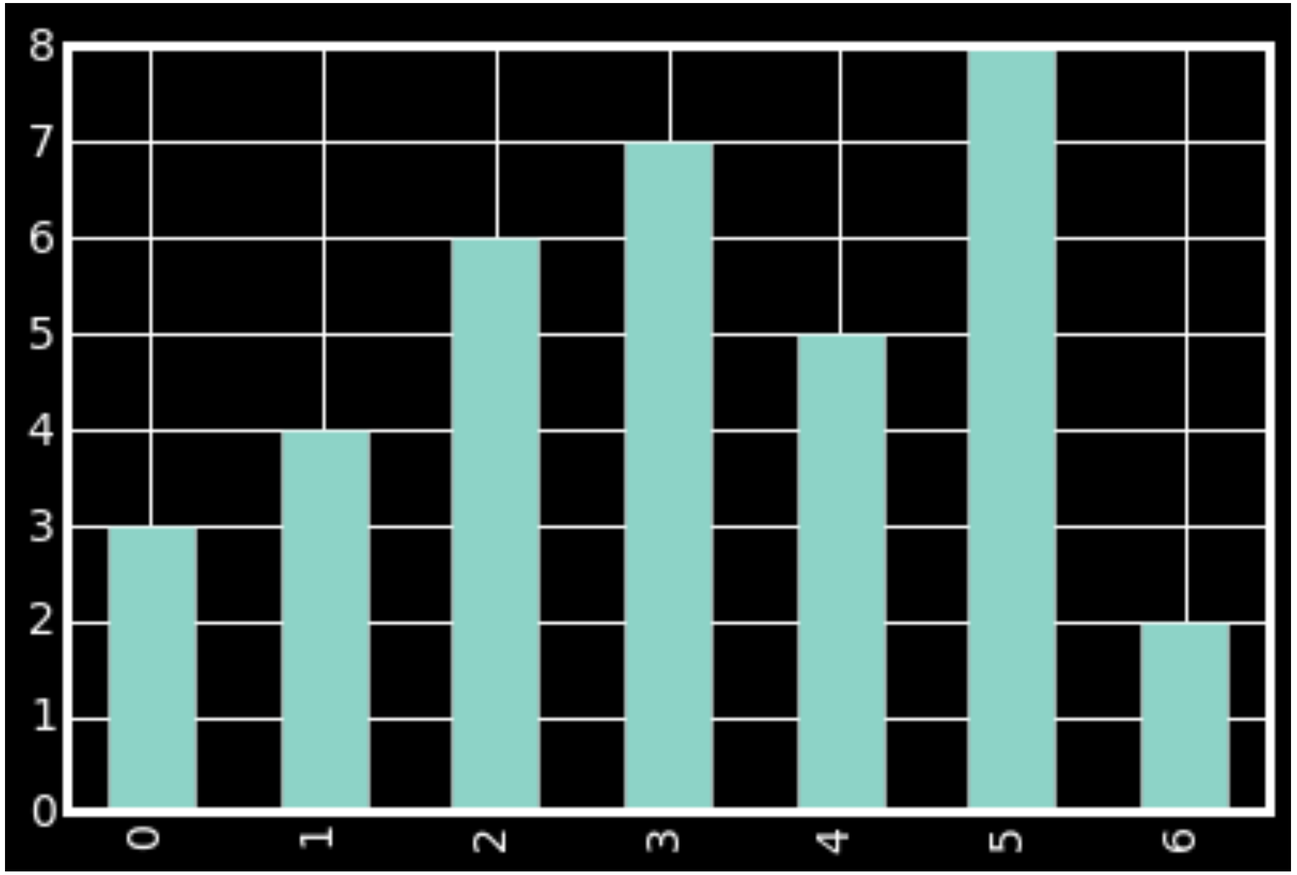
UGLY

METIS



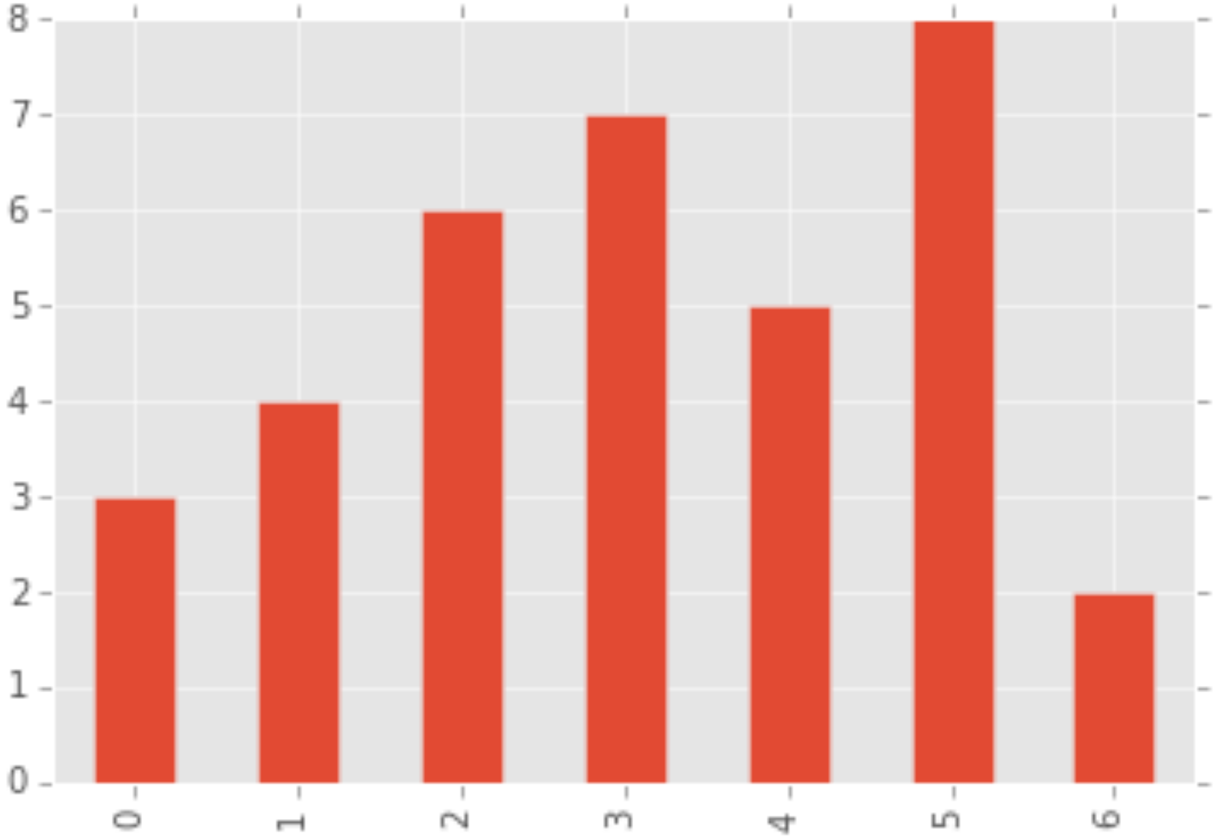

```
data = pd.Series( [3,4,6,7,5,8,2 ])
barchart = data.plot( kind="bar" )
```

METIS



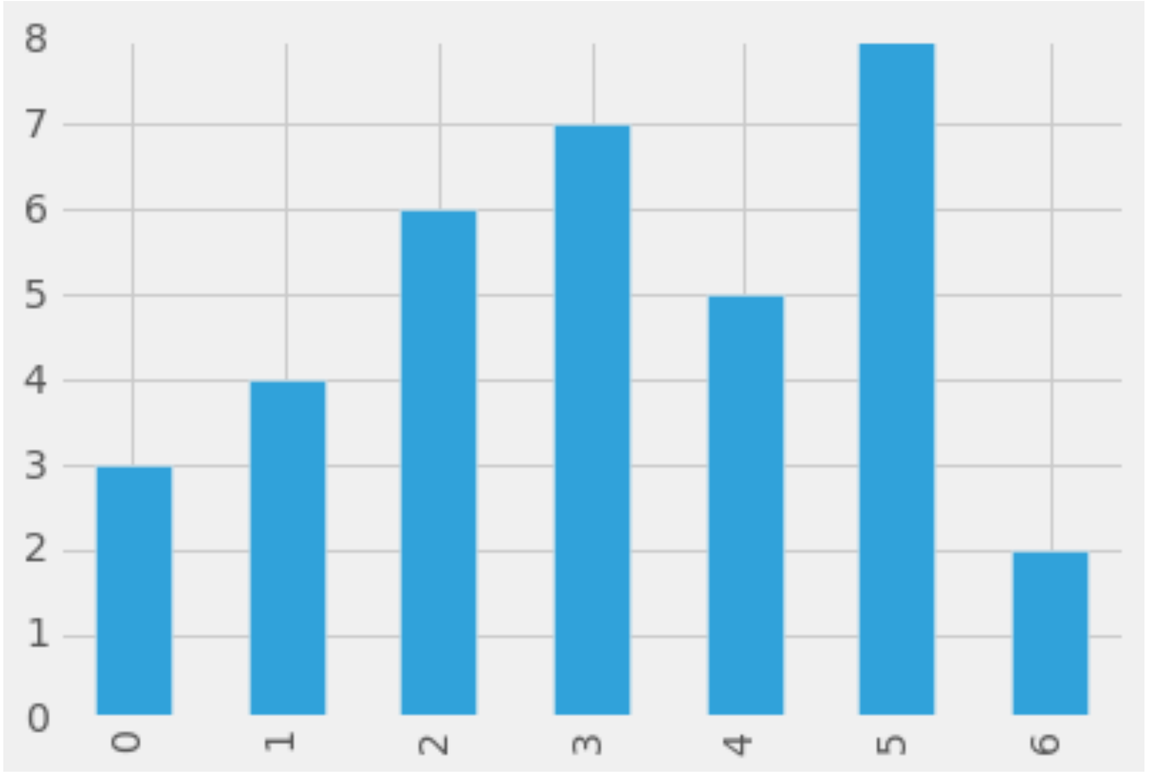

```
plt.style.use('dark_background')
barchart = data.plot( kind="bar" )
```

METIS



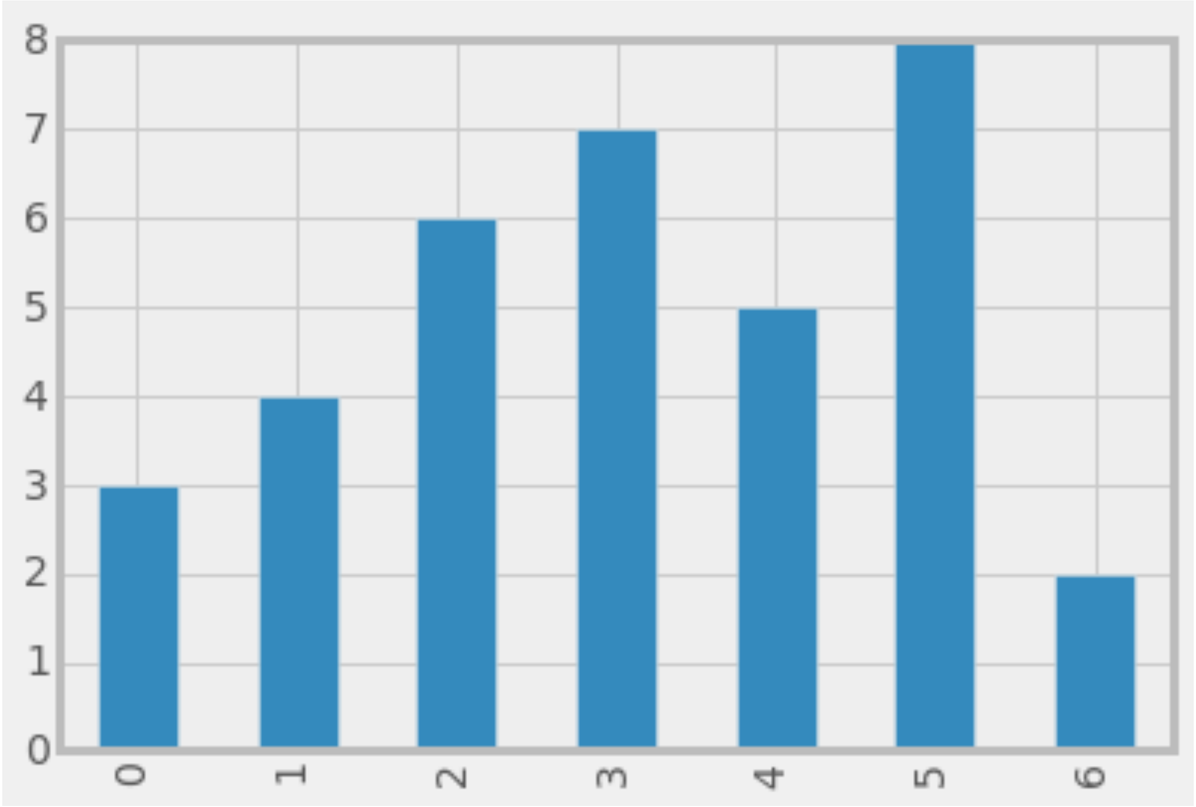

```
plt.style.use('ggplot')
barchart = data.plot( kind="bar" )
```

METIS

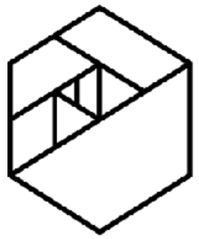



```
plt.style.use('fivethirtyeight')
barchart = data.plot( kind="bar" )
```

METIS

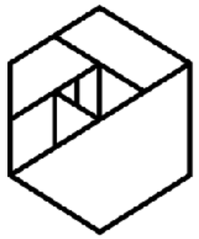



```
plt.style.use('bmh')
barchart = data.plot( kind="bar" )
```



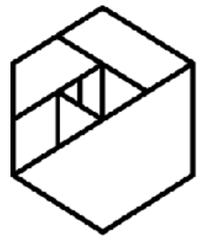
METIS

More Advanced Visualizations



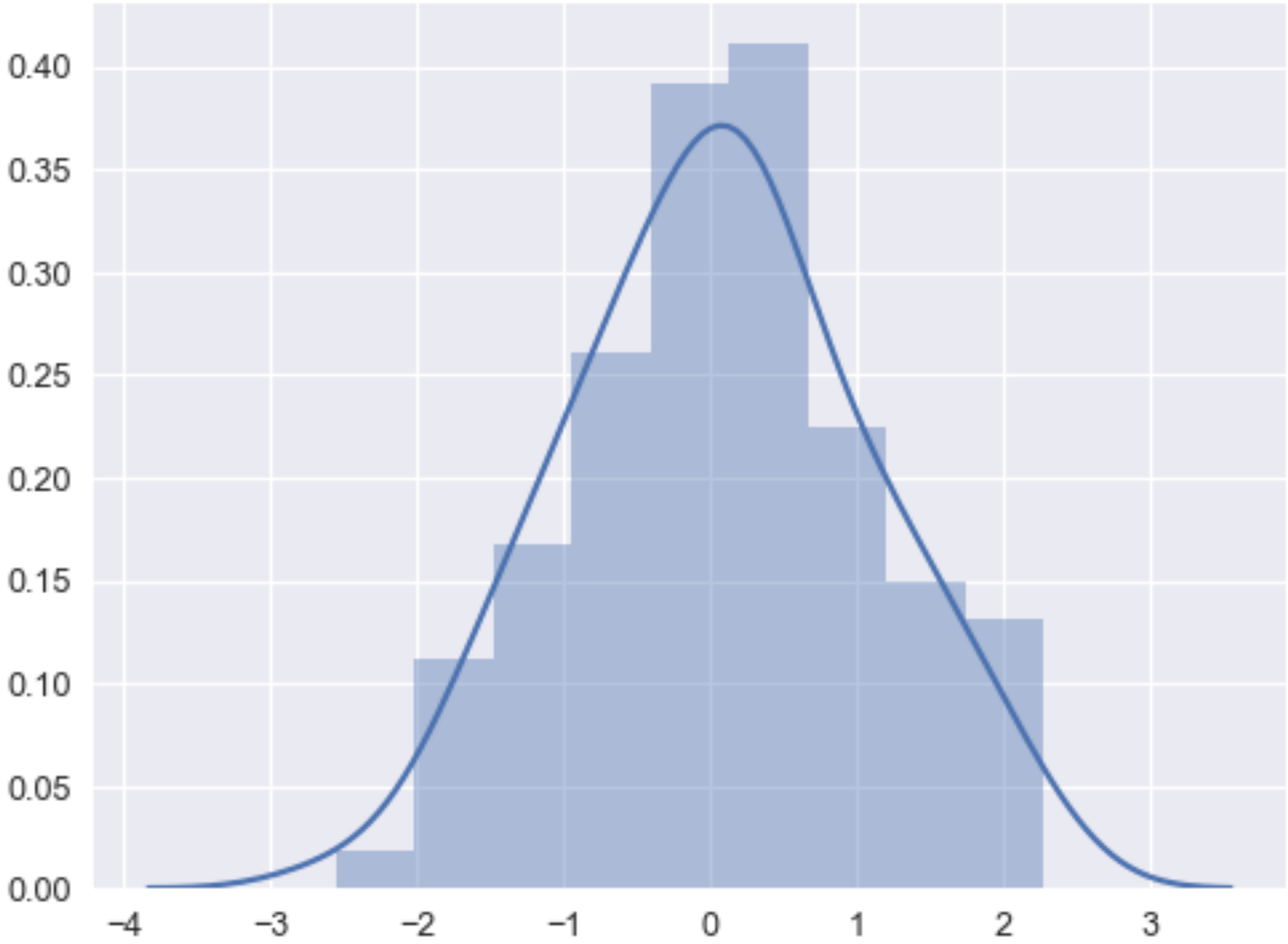
METIS

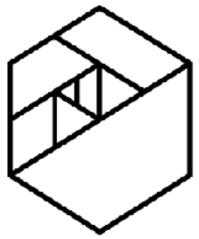
```
import seaborn as sns
```



METIS

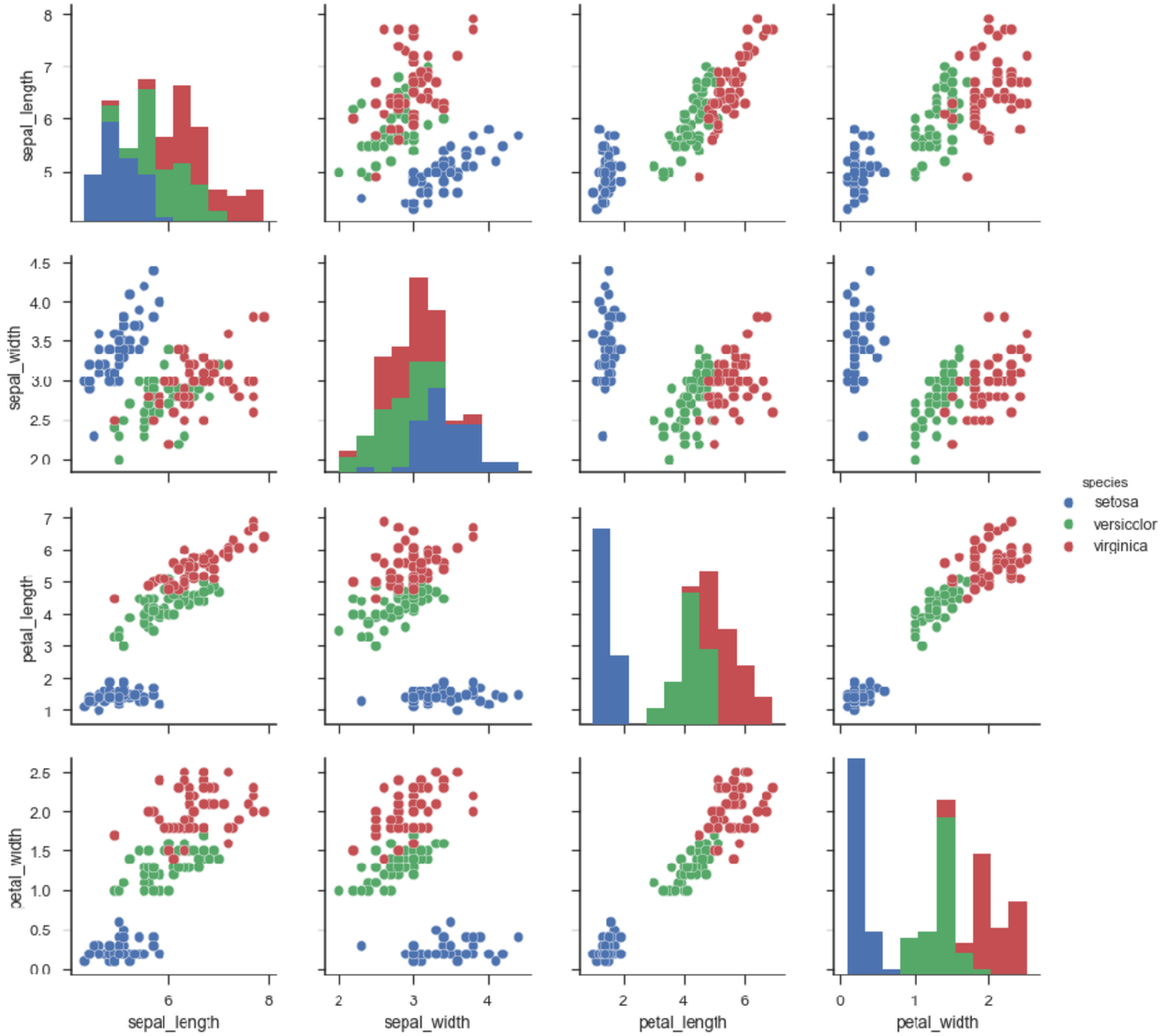
```
ax = sns.distplot(<data>)
```

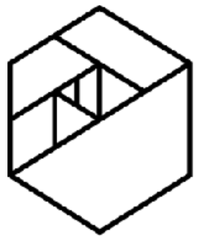




METIS

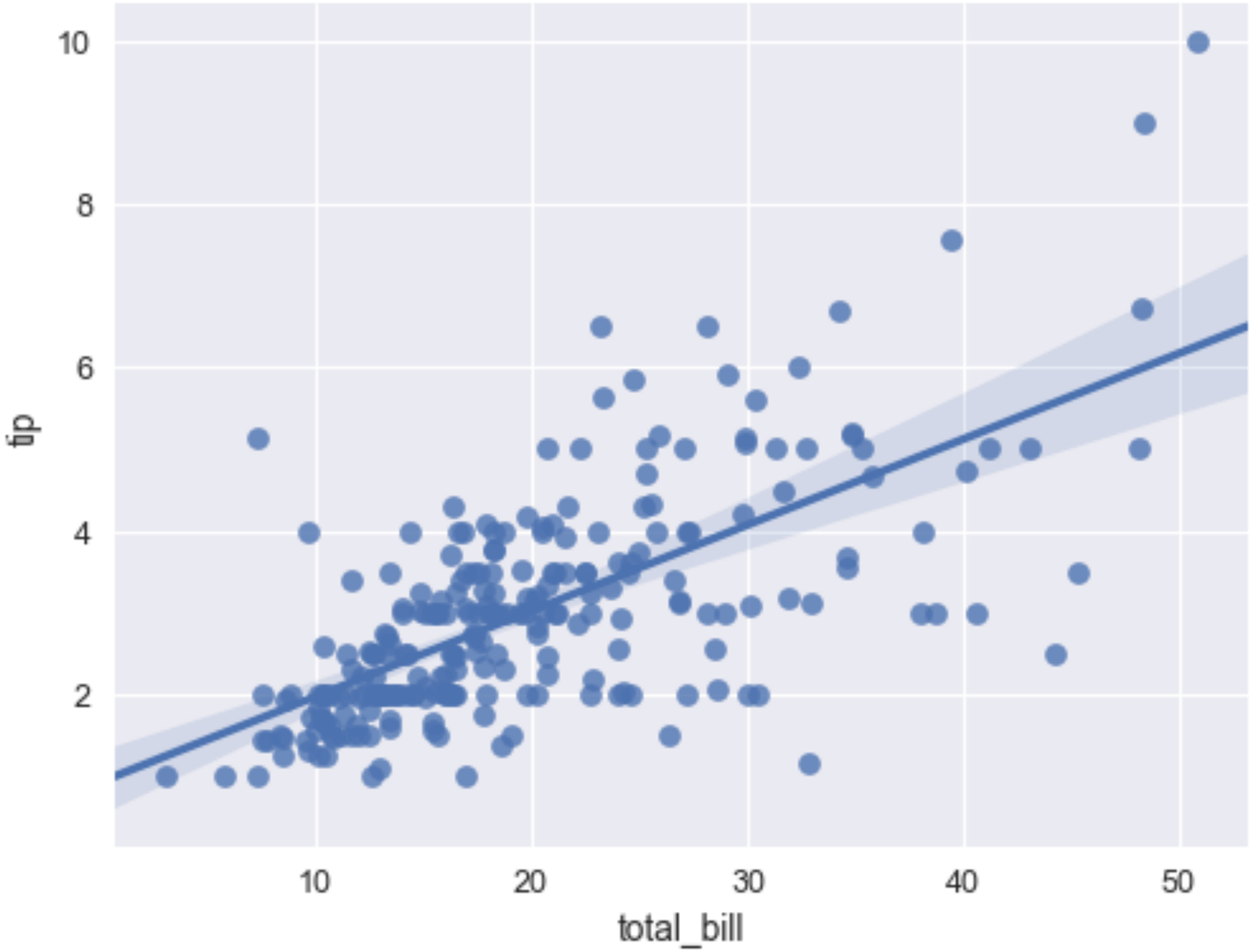
```
g = sns.pairplot(<data>,
    hue="<target>")
```

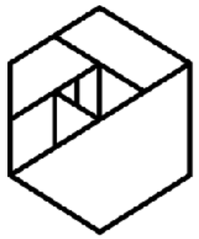




METIS

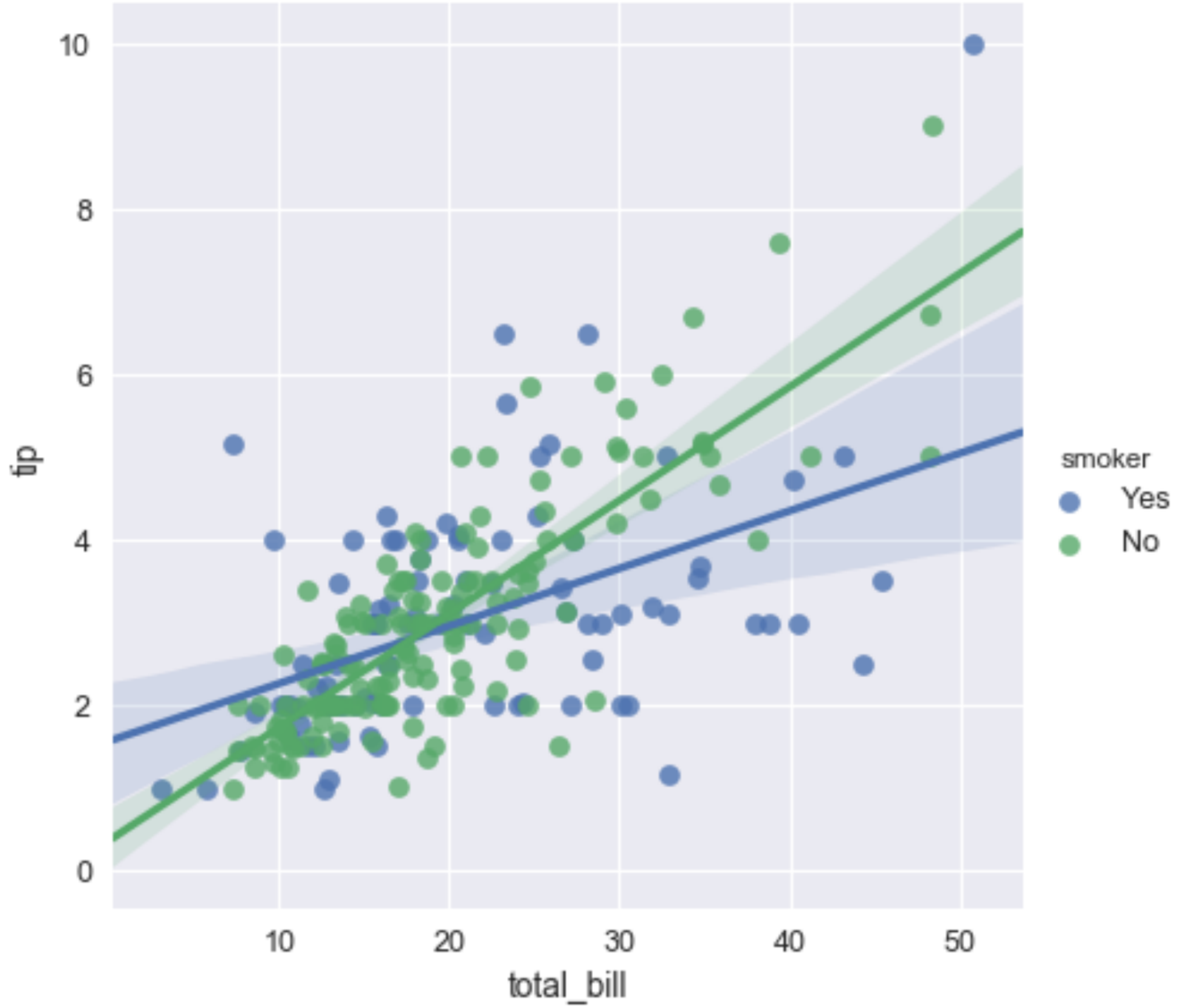
```
ax = sns.regplot(x="total_bill",
                 y="tip", data=tips)
```

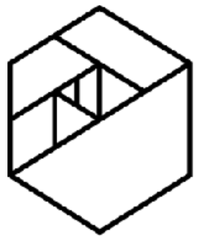




METIS

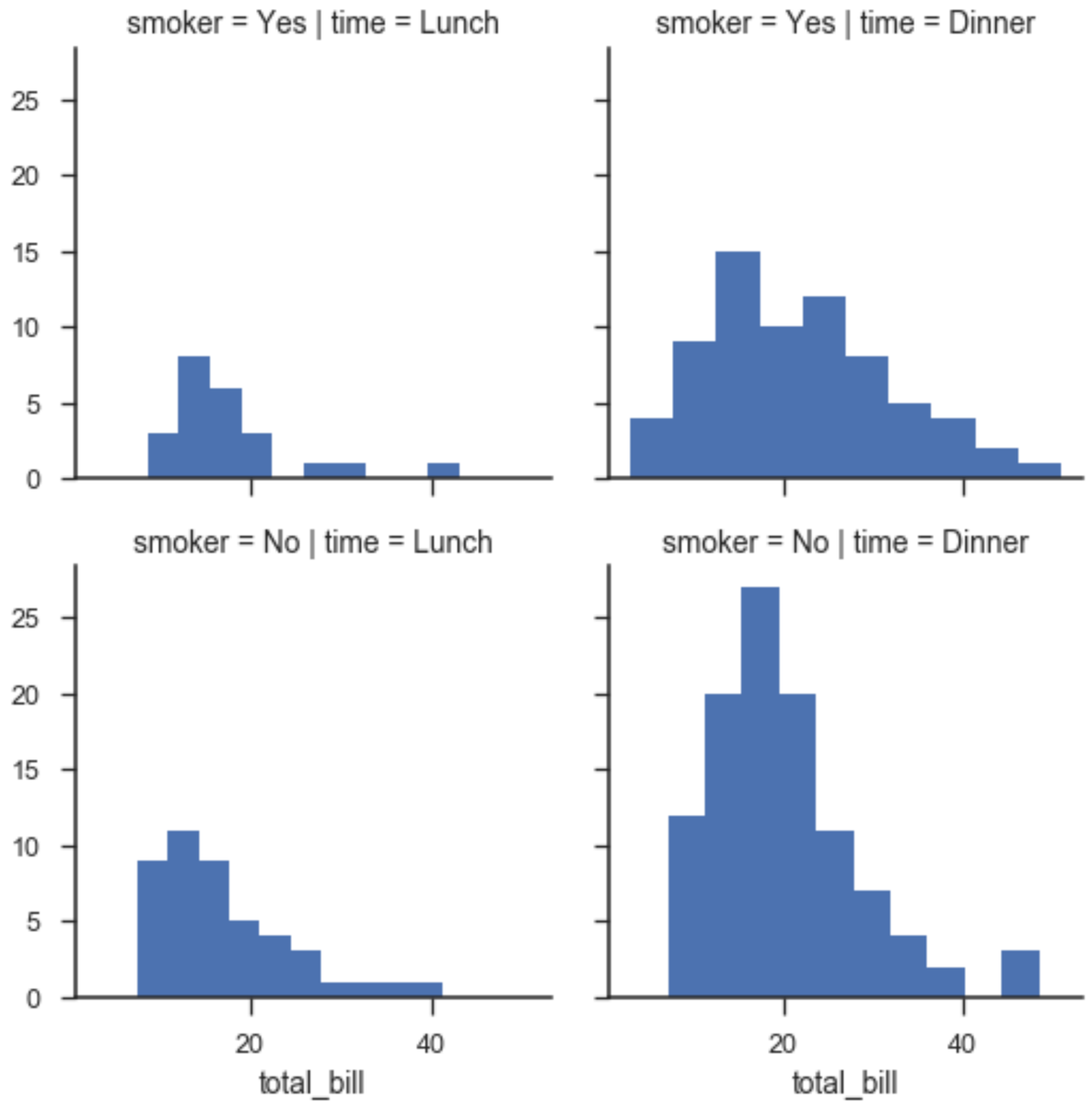
```
g = sns.lmplot(x="total_bill",
y="tip", hue="smoker", data=tips)
```

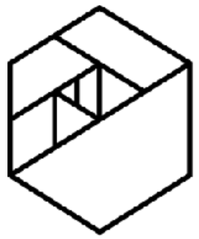




METIS

```
g = sns.FacetGrid(tips, col="time",  row="smoker")
>>> g = g.map(plt.hist, "total_bill")
```

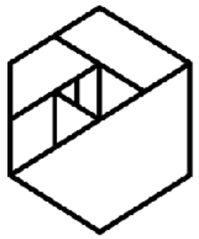




METIS

In Class Exercise

Homework Exercises 1-3



METIS

Questions?