Statistical Visualization

Nancy Reid
University of Toronto

HARVARD UNIVERSITY

Celebrating the Foundations and Impact of Statistics: A Symposium Honoring the 95th Birthday of Herman Chernoff

Background

- I'm an amateur
- But I have professional colleagues

Fanny Chevalier

U Toronto

Sheelagh Carpendale

U Calgary

Chris Collins

UOIT

Background

- And I organized a workshop

JANUARY 12 – 23, 2015

Opening Conference and Boot Camp

Organizing Committee: Nancy Reid (Chair), Sallie Keller, Lisa Lix, Bin Yu

JANUARY 26 – 30, 2015

Workshop on Big Data and Statistical Machine Learning

Organizing committee: Ruslan Salakhutdinov (Chair), Dale Schuurmans, Yoshua Beng

Hugh Chipman, Bin Yu

FEBRUARY 9 – 13 , 2015

Workshop on Optimization and Matrix Methods in Big Data

Organizing Committee: Stephen Vavasis (Chair), Anima Anandkumar, Petros Drineas

Michael Friedlander , Nancy Reid, Martin Wainwright

FEBRUARY 23 – 27, 2015

Workshop on Visualization for Big Data: Strategies and Principles

Organizing Committee: Nancy Reid (Chair), Susan Holmes, Snehelata Huzurbazar, Hadley Wickham, Leland Wilkinson

MARCH 23 – 27, 2015

Workshop on Big Data in Health Policy

Organizing Committee: Lisa Lix (Chair), Constantine Gatsonis , Sharon-Lise Norman

Background

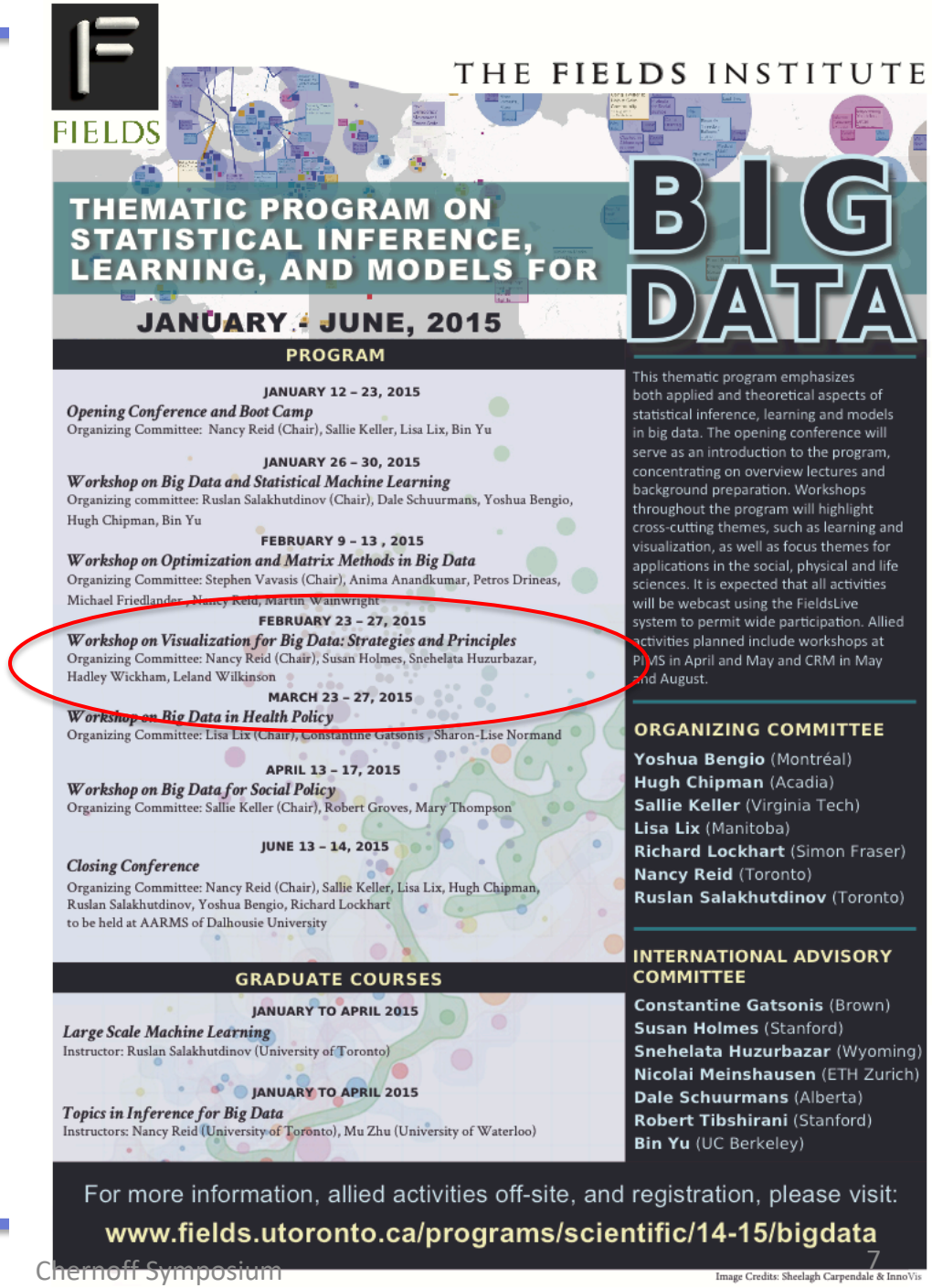
- And I organized a workshop

Background

- And I organized a workshop
- And I went to all the talks

Background

- And I organized a workshop
- And I went to all the talks
- Which were great



Background

- And I organized a workshop
- And I went to all the talks
- Which were great
- And all online

Background

- And I organized a workshop
- And I went to all the talks
- Which were great
- And all online

FieldsLive

Background

- Statistical graphics
- Information visualization
- Scientific visualization
- Journalism

2018

Me to Fanny –

"are Chernoff faces a thing now?"

Fanny Chevalier

U Toronto

2018

Me to Fanny –

"are Chernoff faces a thing now?"

Fanny to me –

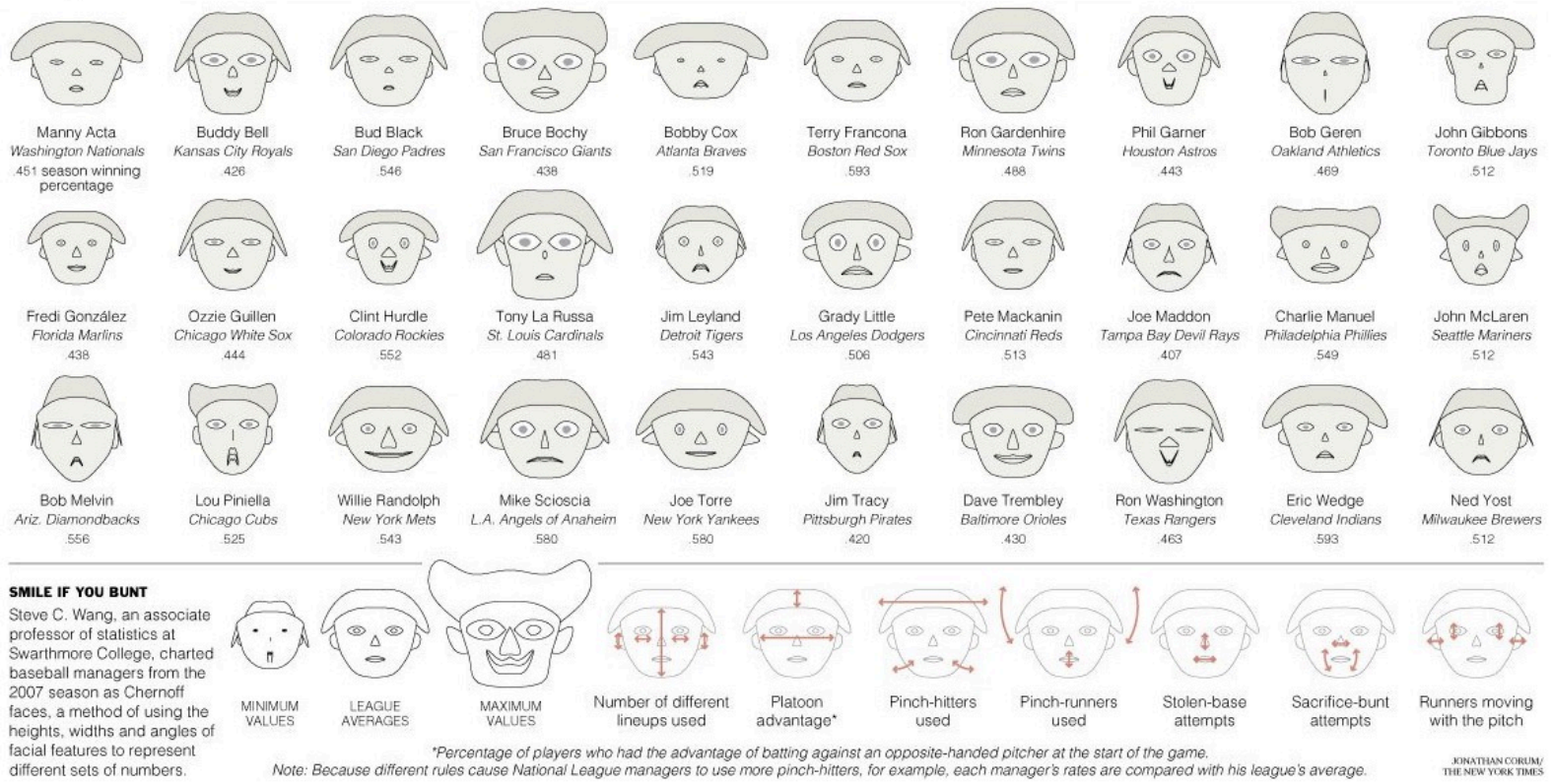
"a great seminal example of rich glyph-based visualization"

Fanny Chevalier

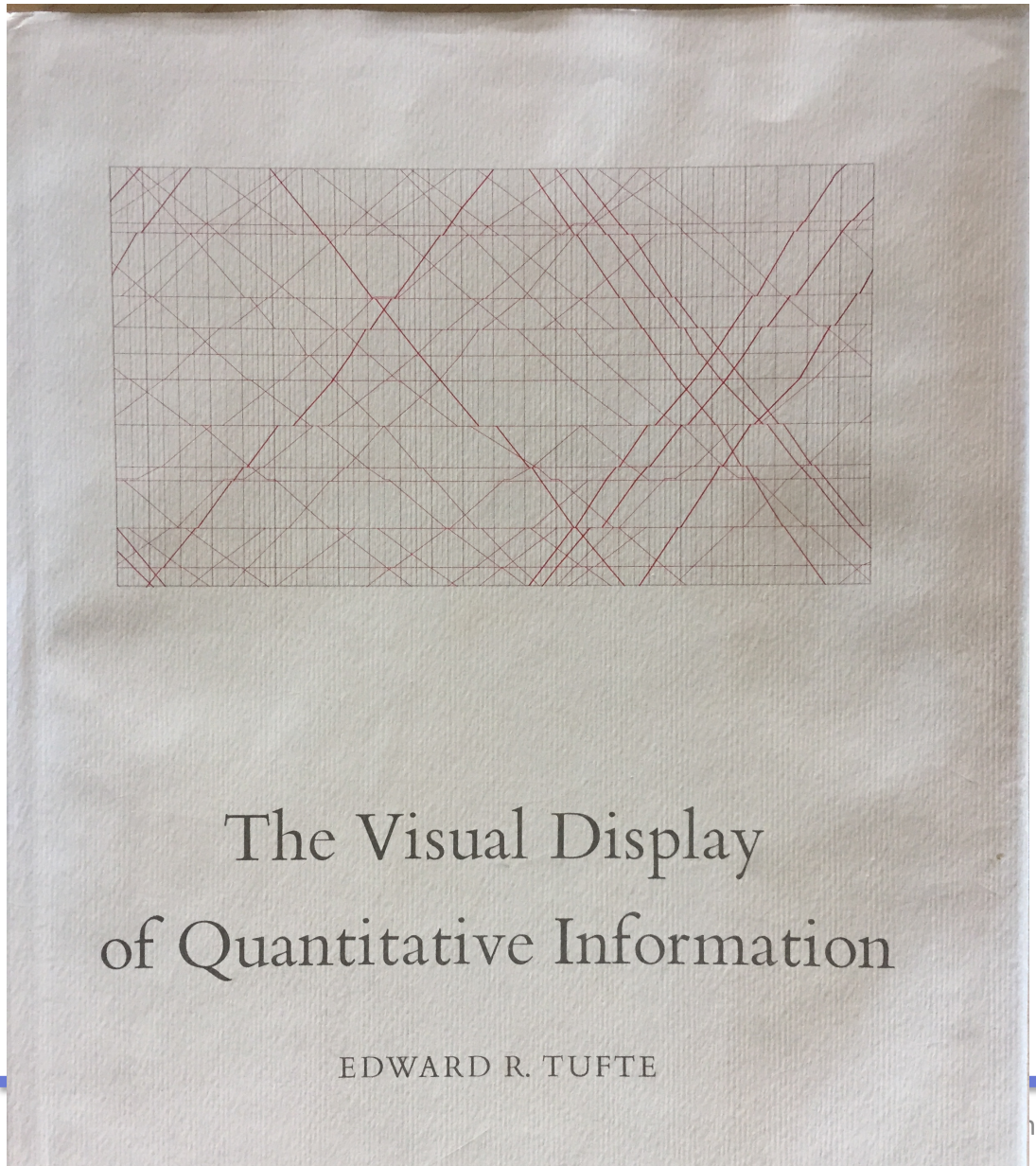
U Toronto

2008 The New York Times "Smile if you bunt"

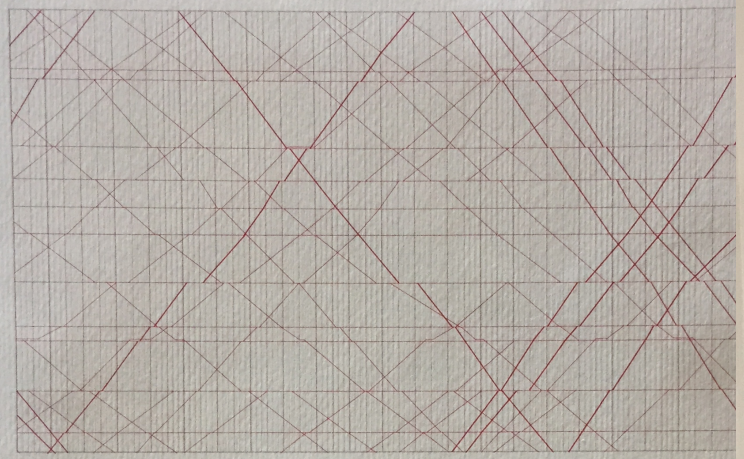
The New York Times April 1, 2008



1983


The Visual Display
of Quantitative Information
EDWARD R. TUFTE

1983

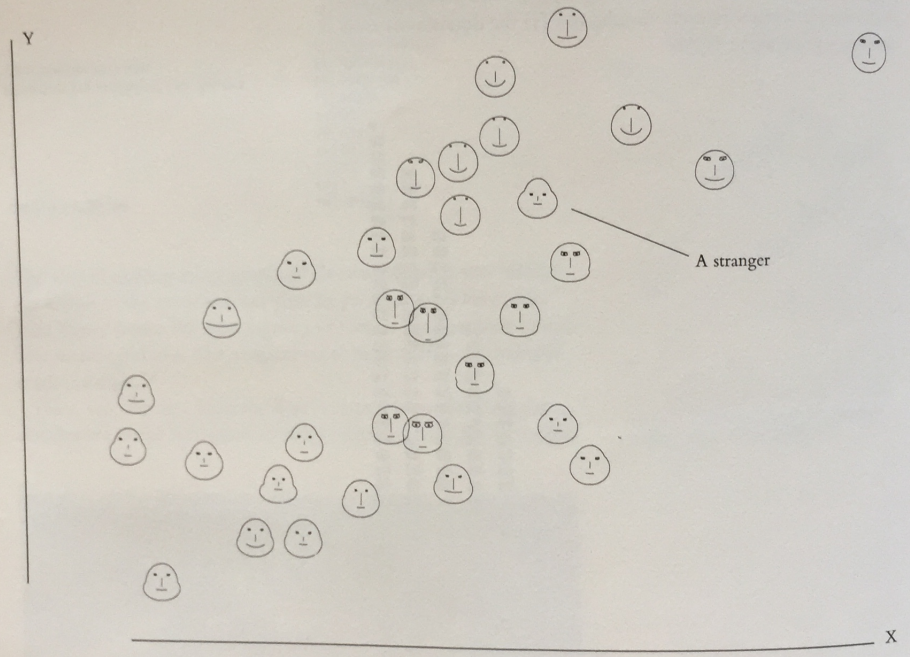


The Visual Display of Quantitative Informat

EDWARD R. TUFTE

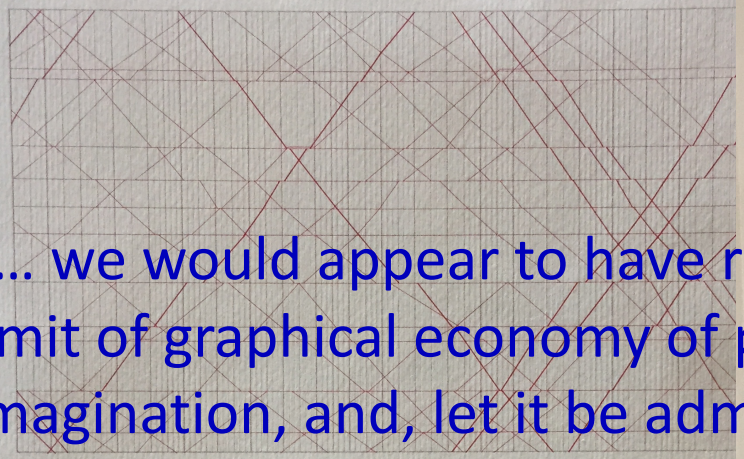
Encoding of data measures can be far more elaborate. The plotted points here are Chernoff faces, which reduce well, maintaining legibility even with individual areas of .05 square inches as shown.[3] The analyst would observe the standard X-Y scatterplot and then turn to the within-scatter detail, seeking clusters of similar observations over the X-Y plane. Outlying faces and those inconsistent with others in the neighborhood—they are, of course, *strangers*—should be identified by observation number or name.

[3] Herman Chernoff, "The Use of Faces to Represent Points in k-Dimensional Space Graphically," *Journal of the American Statistical Association* 68 (June 1973), 361–368. For an application of faces located over two dimensions, see Howard Wainer and David Thissen, "Graphical Data Analysis," *Annual Review of Psychology*, 32 (1981), 191–241.



With cartoon faces and even numbers becoming data measures,

1983



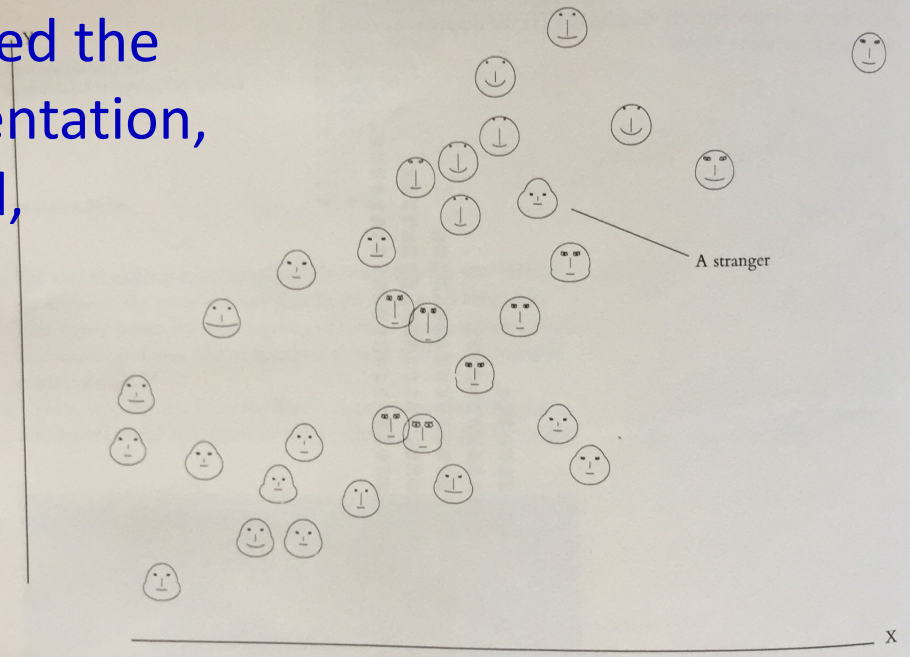
"… we would appear to have reached the limit of graphical economy of presentation, imagination, and, let it be admitted, eccentricity"

The Visual Display of Quantitative Informat

EDWARD R. TUFTE

Encoding of data measures can be far more elaborate. The plotted points here are Chernoff faces, which reduce well, maintaining legibility even with individual areas of .05 square inches as shown.[3] The analyst would observe the standard X-Y scatterplot and then turn to the within-scatter detail, seeking clusters of similar observations over the X-Y plane. Outlying faces and those inconsistent with others in the neighborhood—they are, of course, *strangers*—should be identified by observation number or name.

[3] Herman Chernoff, "The Use of Faces to Represent Points in k-Dimensional Space Graphically," *Journal of the American Statistical Association* 68 (June 1973), 361–368. For an application of faces located over two dimensions, see Howard Wainer and David Thissen, "Graphical Data Analysis," *Annual Review of Psychology*, 32 (1981), 191–241.



With cartoon faces and even numbers becoming data measures,

1972

The Use of Faces to Represent Points in k-Dimensional Space Graphically

HERMAN CHERNOFF *

A novel method of representing multivariate data is presented. Each point in k-dimensional space, $k \leq 18$, is represented by a cartoon of a face whose features, such as length of nose and curvature of mouth, correspond to components of the point. Thus every multivariate observation is visualized as a computer-drawn face. This presentation makes it easy for the human mind to grasp many of the essential regularities and irregularities present in the data. Other graphical representations are described briefly.

1. INTRODUCTION

A graphical method of representing points in k-dimensional space is presented for consideration. The object is to represent multivariate data, subject to strong but possibly complex relationships, in such a way that an investigator can quickly comprehend relevant information and then apply appropriate statistical analysis. The method consists of representing a point in k-dimensional space by a drawing of a face whose characteristics are determined by the position of the point. A sample of

2.1 Fossil Data

Eight measurements were made on each of 88 nummulited specimens from the Eocene Yellow Limestone Formation of northwestern Jamaica. Two measurements thought to be age-dependent were discarded. One specimen (Number 34) was rejected because of a permutation in an early copy of the measurements for that specimen which cast doubt upon its accuracy. The data for Example 1 (see Table 1′ in the appendix) represent

Z_1 inner diameter of embryonic chamber (in microns)
Z_2 total number of whorls
Z_3 number of chambers in first whorl
Z_4 number of chambers in last whorl
Z_5 maximum height of chambers in first whorl (in microns)
Z_6 maximum height of chambers in last whorl (in microns)

The 87 faces corresponding to the 87 remaining specimens are presented in sequential order as indicated in Figure 1A. This order was selected *after* the data had been grouped into three clusters by an analytic technique [9].

1972

1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

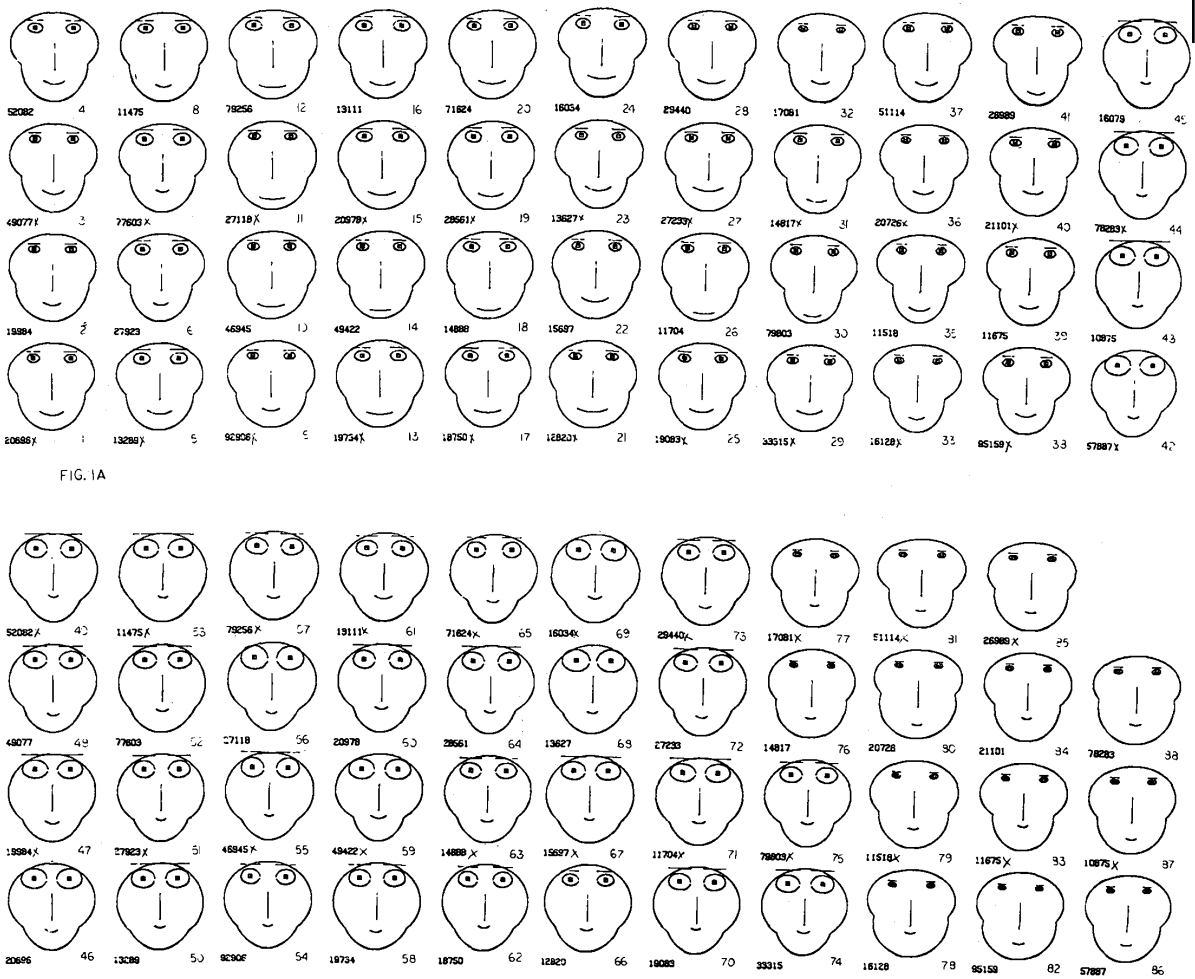


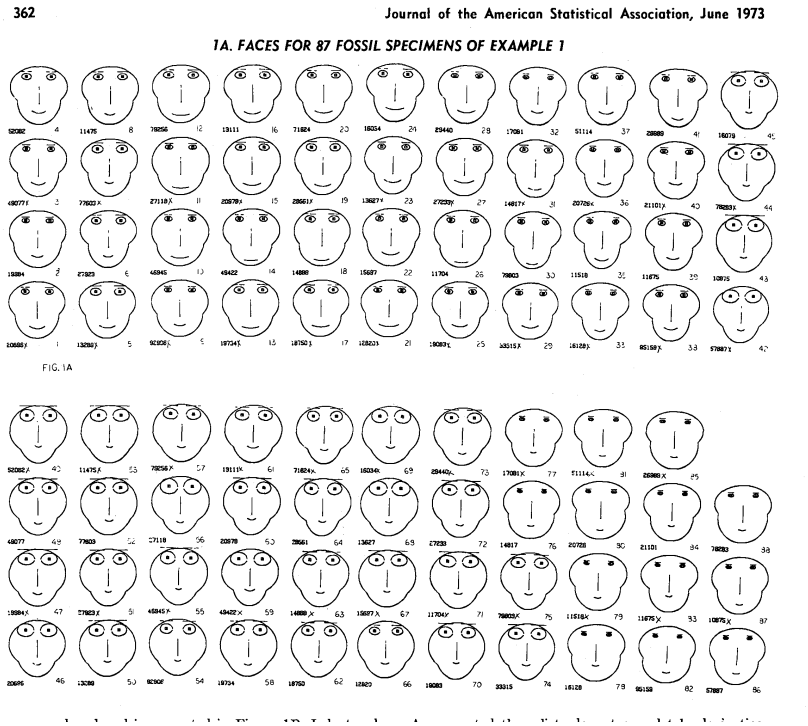
FIG. 1A

was produced and is presented in Figure 1B. I clustered these visually. The groups were:

I: (1, 2, 3, 9, 22, 29)
II: (4, 5, 6, 7, 8)
III: (10, 11, 14, 23, 25, 26, 27)

As presented these lists do not completely do justice to the information obtained from the faces. They do not indicate which specimens are obviously members of a group and which were regarded as borderline. Each specimen was forced into some group, whereas one would

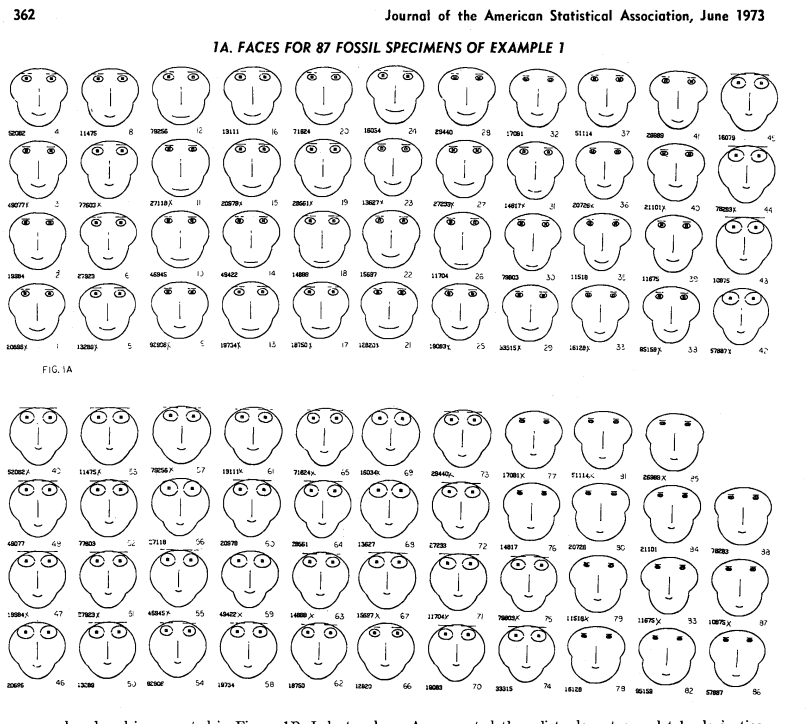
The rise of the machine


362 Journal of the American Statistical Association, June 1973

1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

FIG. 1A

The rise of the machine

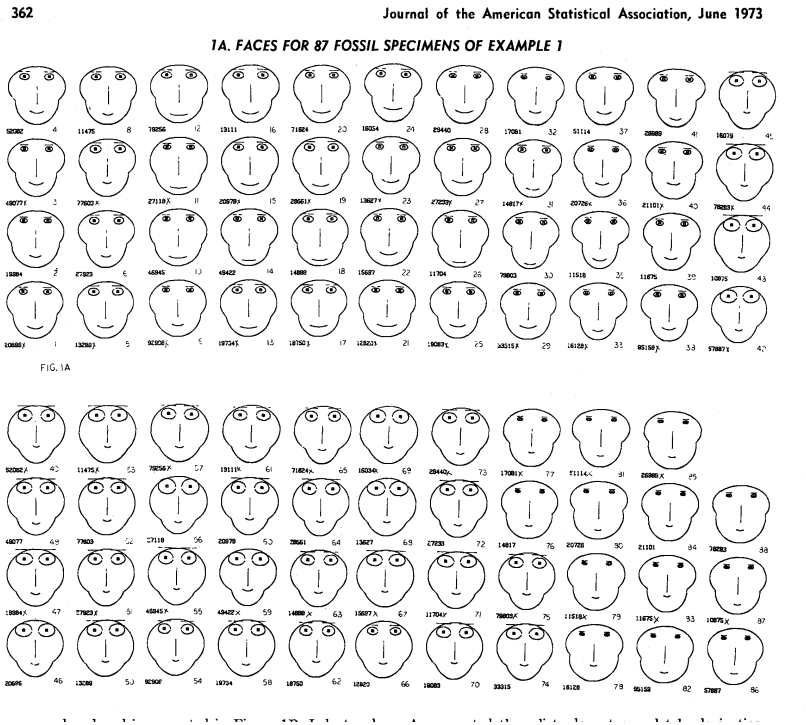

362 Journal of the American Statistical Association, June 1973

1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

FIG. 1A

"At this time the cost of drawing these faces is about 20 to 25 cents per face on the IBM 360-67 at Stanford University using the Calcomp Plotter. Most of this cost is in the computing, and I believe that it should be possible to reduce it considerably"

The rise of the machine

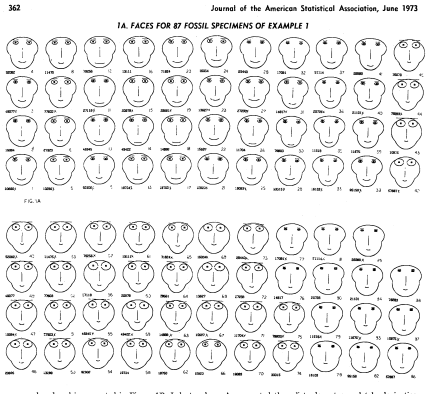

362 Journal of the American Statistical Association, June 1973

1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

FIG. 1A

"At this time the cost of drawing these faces is about 20 to 25 cents per face on the IBM 360-67 at Stanford University using the Calcomp Plotter. Most of this cost is in the computing, and I believe that it should be possible to reduce it considerably"

The rise of AI


362 Journal of the American Statistical Association, June 1973

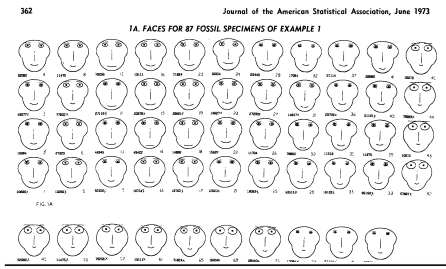
1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

CANADA

The Man Who Helped Turn Toronto Into a High-Tech Hotbed

Geoffrey Everest Hinton, a computer scientist and leading expert in artificial intelligence, has helped make the University of Toronto a center of innovation. Aaron Vincent Elkaim for The New York Times

The rise of AI


362 Journal of the American Statistical Association, June 1973

1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

CANADA

The Man Who Helped Turn Toronto Into a High-Tech Hotbed

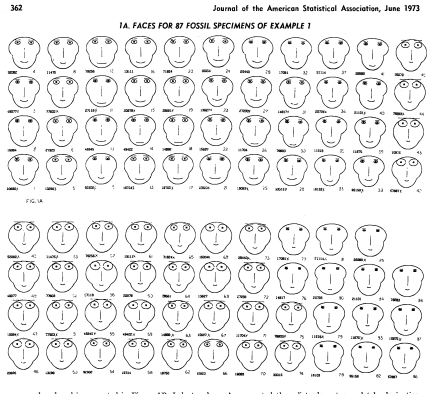
TECHNOLOGY

How Artificial Intelligence Is Edging Its Way Into Our Lives

查看简体中文版 | 查看繁體中文版

By THE NEW YORK TIMES FEB. 12, 2018

The rise of AI

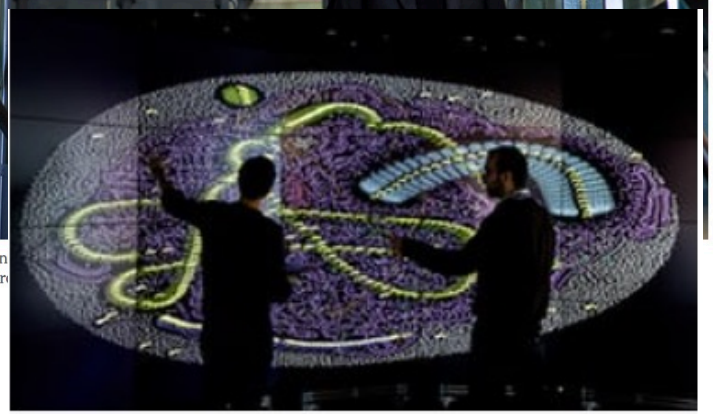

362 Journal of the American Statistical Association, June 1973

1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

CANADA

The Man Who Helped Turn Toronto Into a High-Tech Hotbed

Geoffrey Everest Hin
the University of Tor



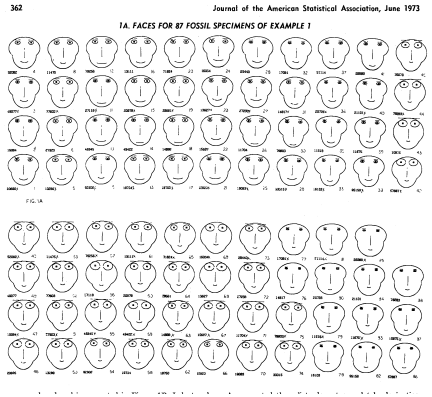
Scientists plan huge European AI hub to compete with US

Exclusive: In an open letter, the scientists say the proposed Ellis institute is essential to avoid brain drain to big tech firms

International edition

The Guardian

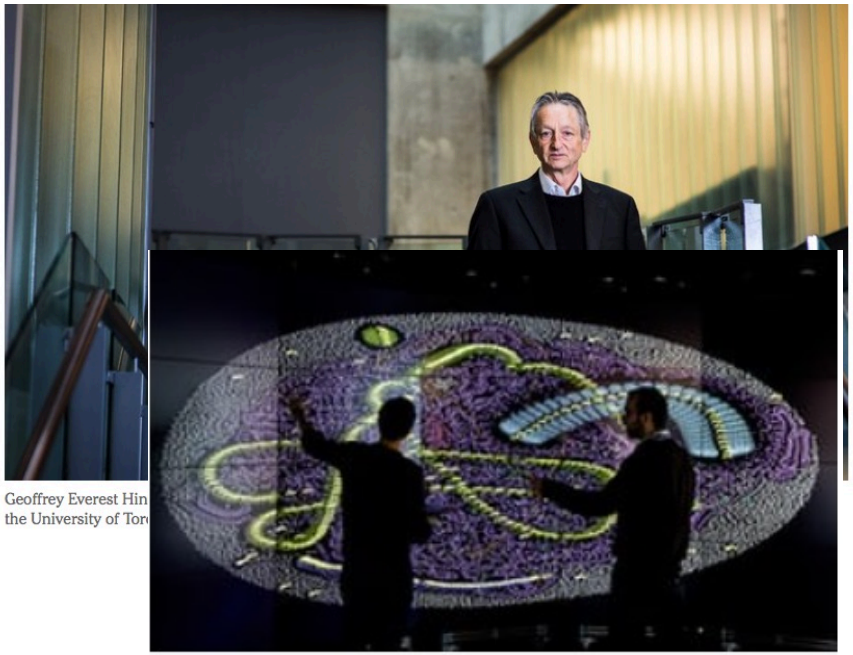
The rise of AI



"This approach is an amusing reversal of a common one in artificial intelligence. Instead of using machines to discriminate between human faces be reducing them numbers, we discriminate between numbers by using the machine to do the brute labor of drawing faces and leaving the intelligence to the humans, who are still more flexible and clever"

CANADA

The Man Who Helped Turn Toronto Into a High-Tech Hotbed



Geoffrey Everest Hin
the University of Tor

Scientists plan huge European AI hub to compete with US

Exclusive: In an open letter, the scientists say the proposed Ellis institute is essential to avoid brain drain to big tech firms

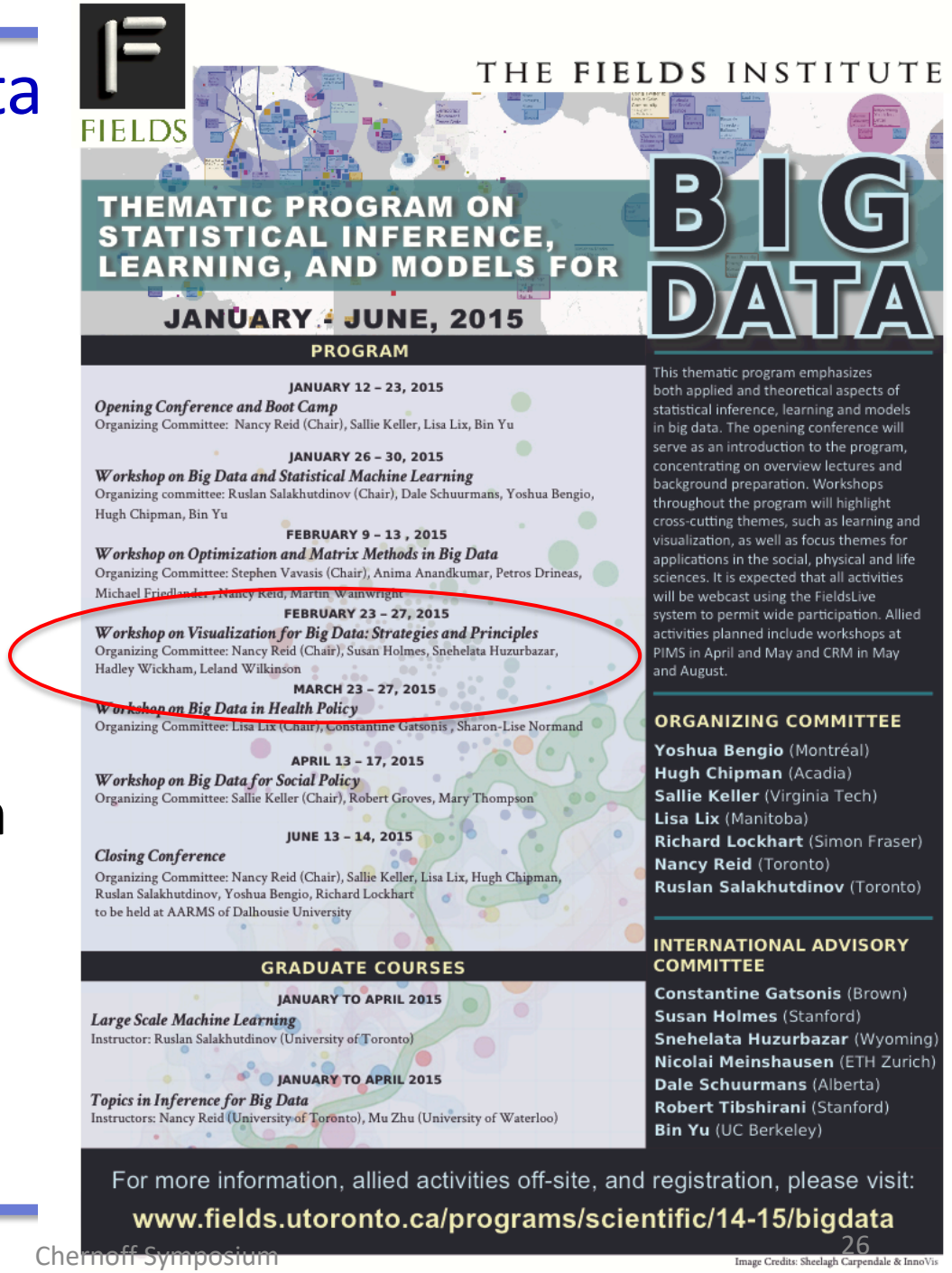
International edition

The Guardian

Visualization for Big Data

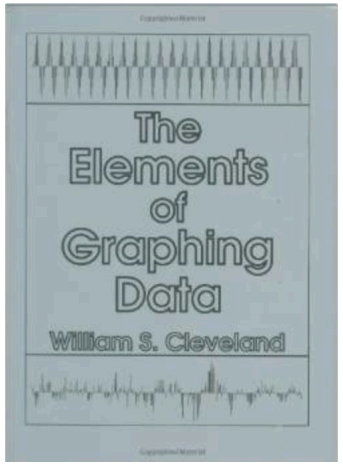
Strategies and Principles

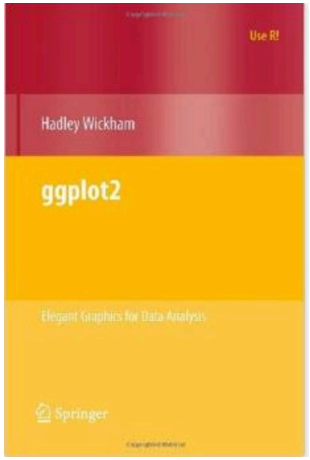
- Statistical graphics
- Information visualization
- Scientific visualization
- Journalism



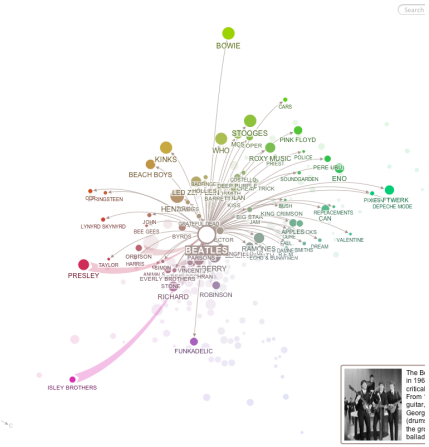
Visualization

- statistical graphics
  - data representation
  - data exploration
  - filtering, sampling aggregation

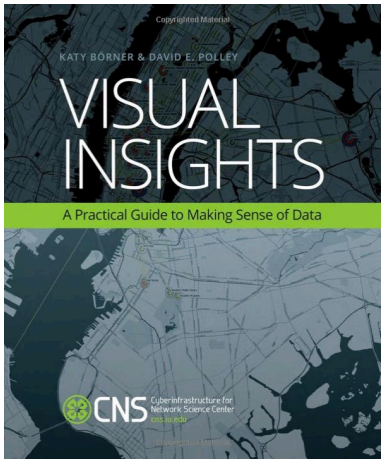




- information visualization



- scientific visualization



- perception, communication, design

Statistical Graphics

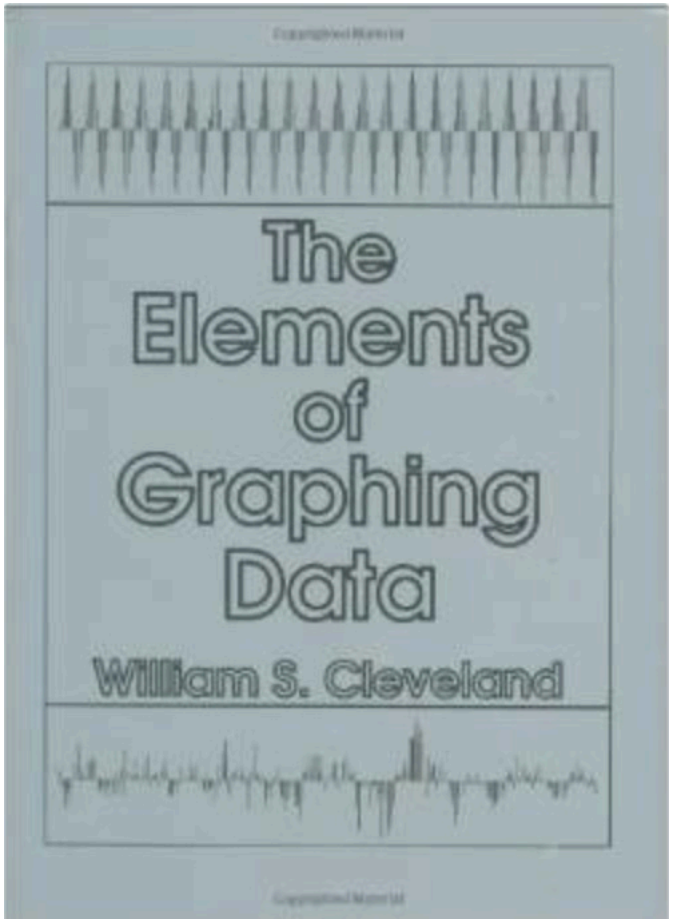
1. Principles of Graph Construction

   terminology, vision, understanding, strategy

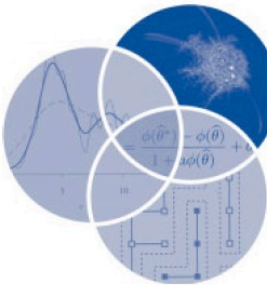
2. Graphical Methods

3. Graphical Perception

   color, texture, scale



Advanced Review



The grammar of graphics

Leland Wilkinson*

The grammar of graphics (GoG) denotes a system with seven classes embedded in a data flow. This data flow specifies a strict order in which data are transformed from a raw dataset to a statistical graphic. Each class contains multiple methods, each of which is a function executed at the step in the data flow corresponding to that class. The classes are orthogonal, in the sense that the product set of all classes (every possible sequence of class methods) defines a space of graphics which is meaningful at every point. The meaning of a statistical graphic is thus determined by the mapping produced by the function chain linking data and graphic. *WIREs Comp Stat* 2010 2 673–677 DOI: 10.1002/wics.118

Keywords: visualization; statistical graphics

Object-oriented graphics system providing a description of the construction

The Grammar of Graphics

1. Specification

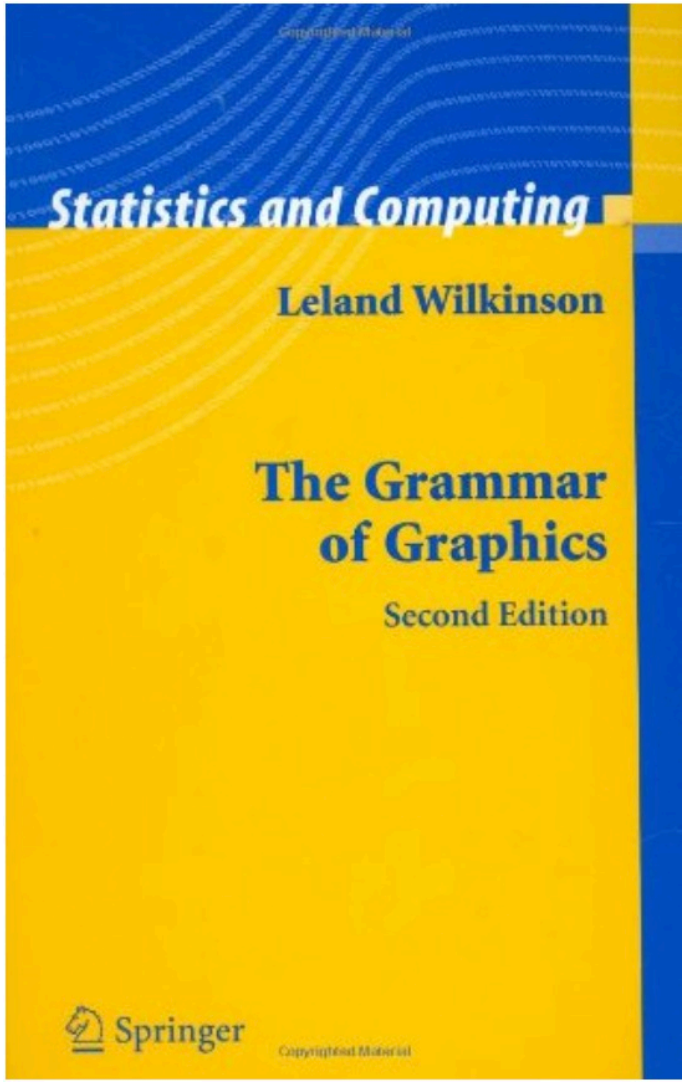
   Data, trans, scale, coord, element, guide

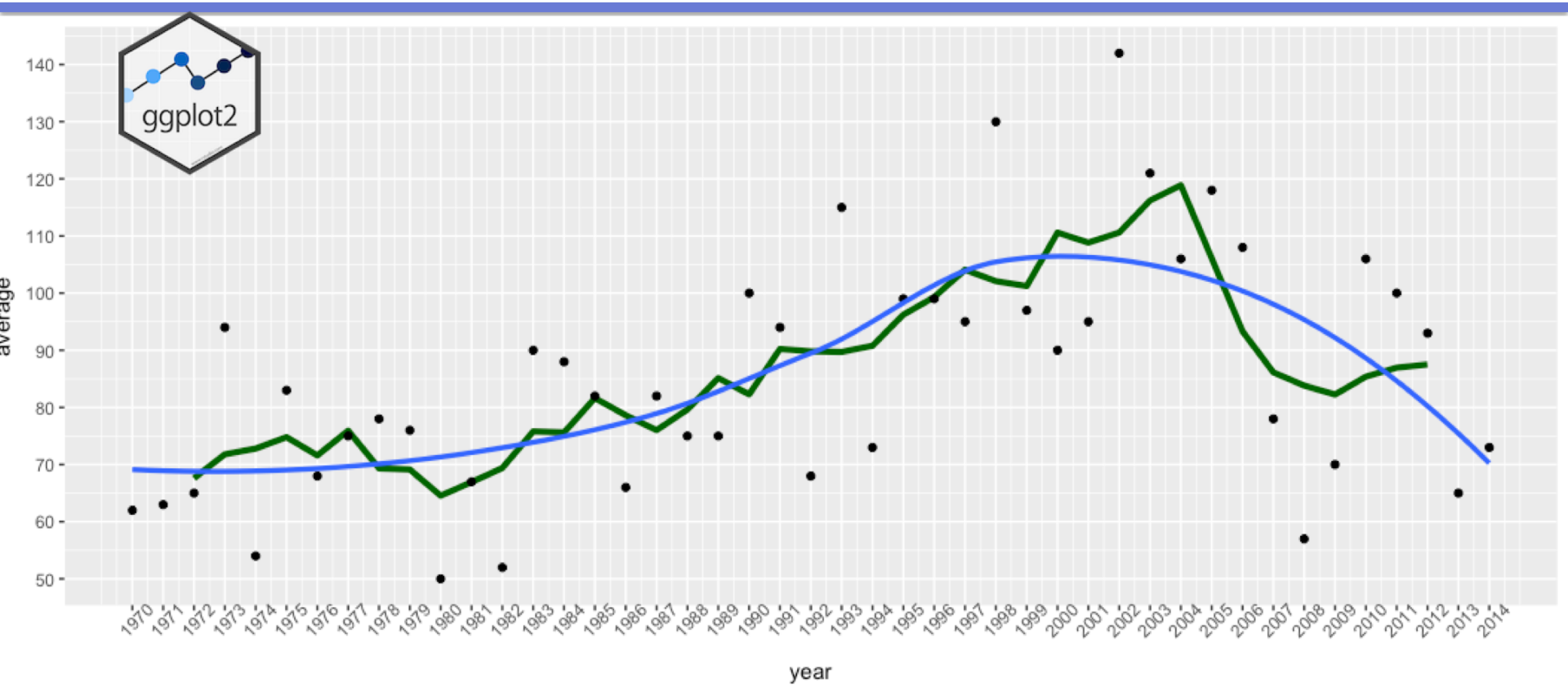
2. Assembly

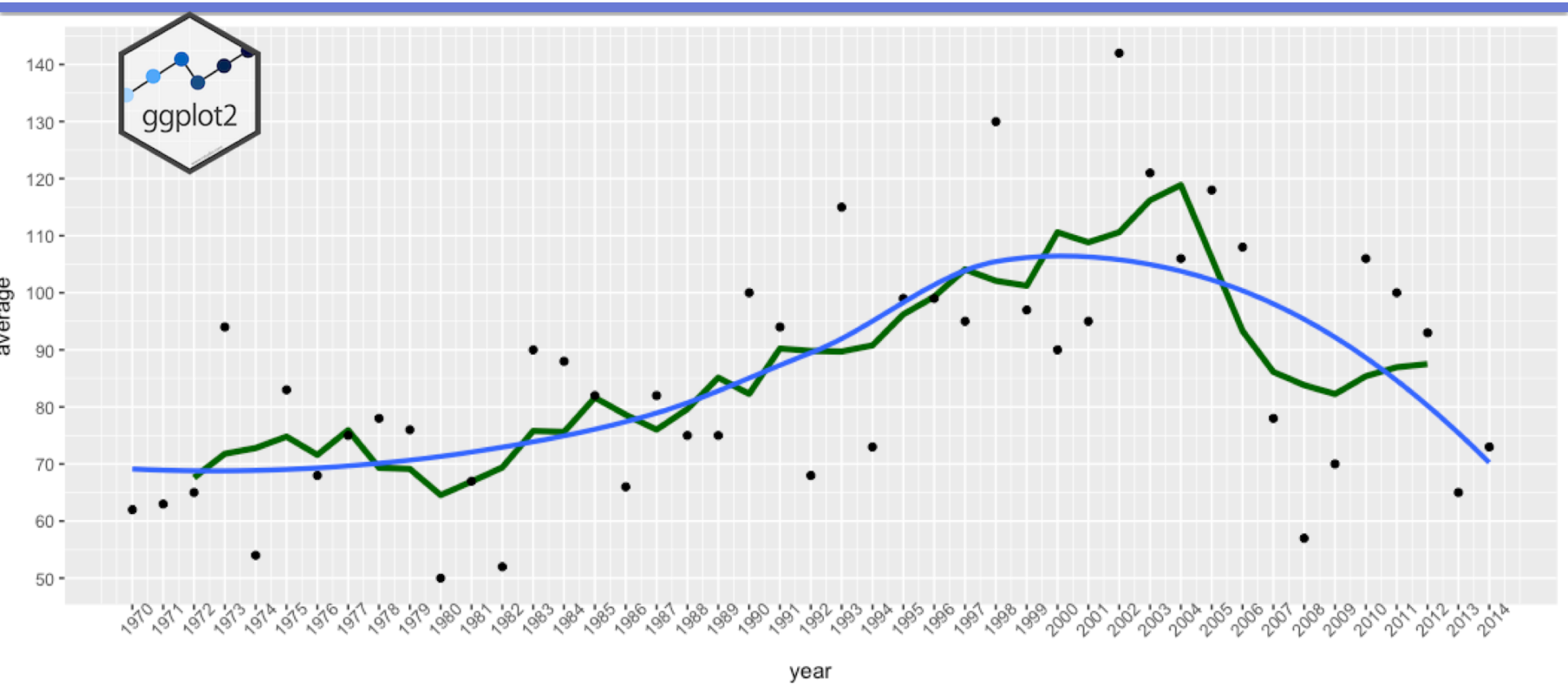
   Geometry, layout, aesthetics

3. Display

   Rendering – static, dynamic, …




ggplot2
average
year
140
130
120
110
100
90
80
70
60
50
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014




```
honeyplot +
   geom_line(aes(honey$year,honey$runmean),col ="green") +
   geom_point(aes(honey$year,honey$average),) +
   scale_x_continuous(breaks=1970:2014) +
   geom_smooth(method="loess",span=.75,se=F) +
   scale_y_continuous(breaks=seq(0,140,by=10)) +
   theme(axis.text.x = element_text(angle=45))
```

The Grammar of Graphics

- `ggplot2 (tidyverse)`

https://github.com/tidyverse/ggplot2

cheatsheets !!

- Tableau visualization system
  - Evolved from Polaris, nViZn

- D3.js
  - Evolved from Protovis

Information Visualization

- a process of transforming information
  into visual form

- relies on the visual system to perceive
  and process the information

- involves the design of visual data representations and interaction techniques

http://www.infovis.org http://ieeevis.org/

Information Visualization

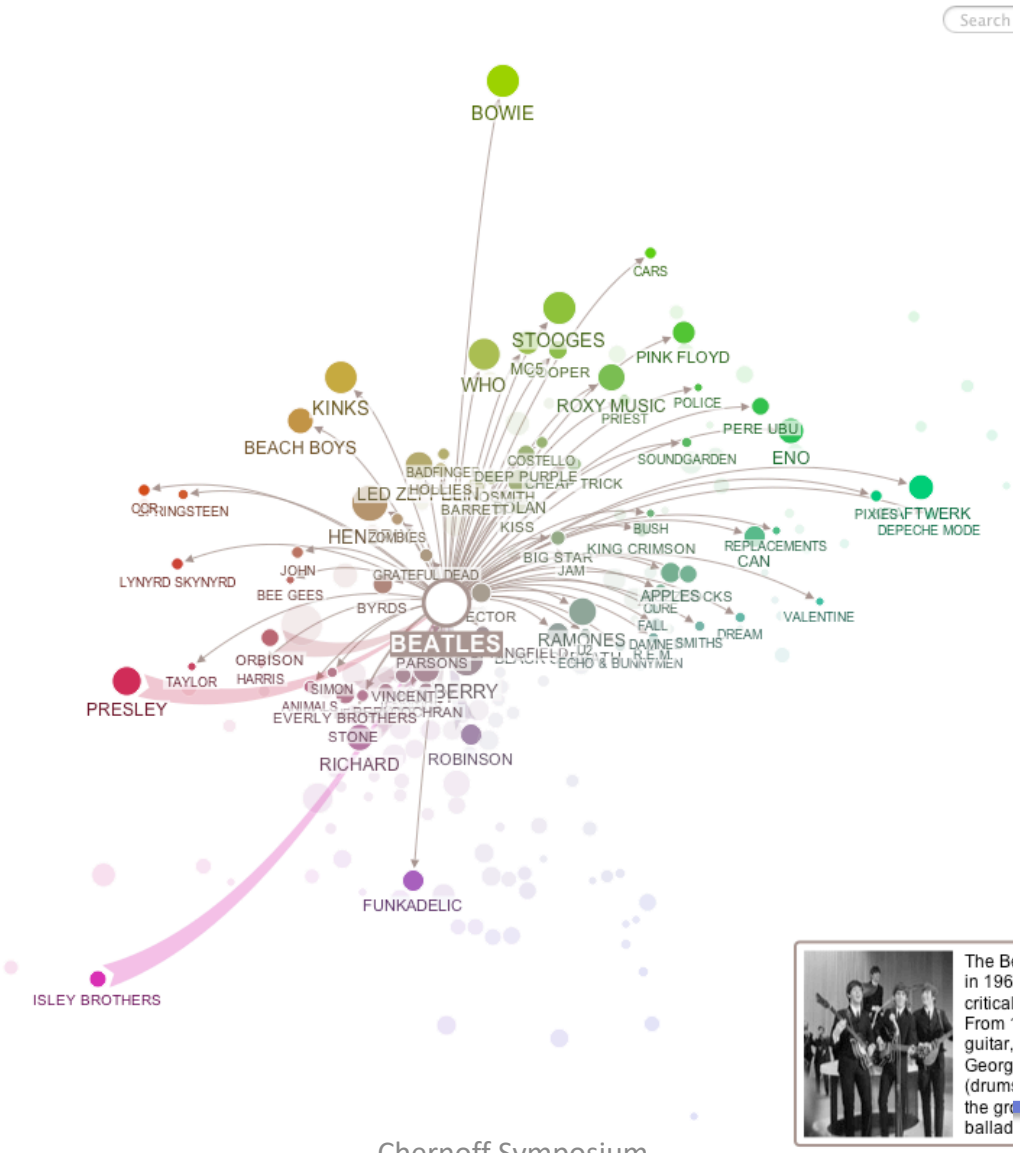
- Representation, presentation, interaction
- “use of computer-supported, interactive, visual representations of data to amplify cognition”

http://innovis.cpsc.ucalgary.ca/

Sheelagh Carpendale, U Calgary

Edge maps

Innovis.cpsc.ucalgary.ca



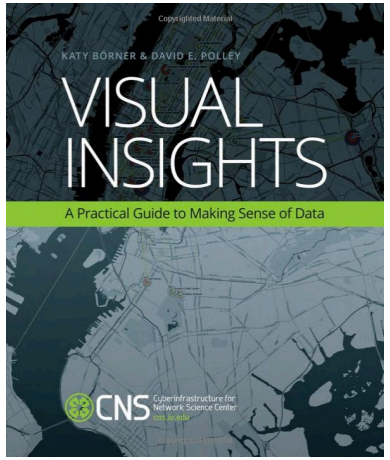
Scientific Visualization

- Katy Borner

- advances understanding or provides solutions for real-world problems

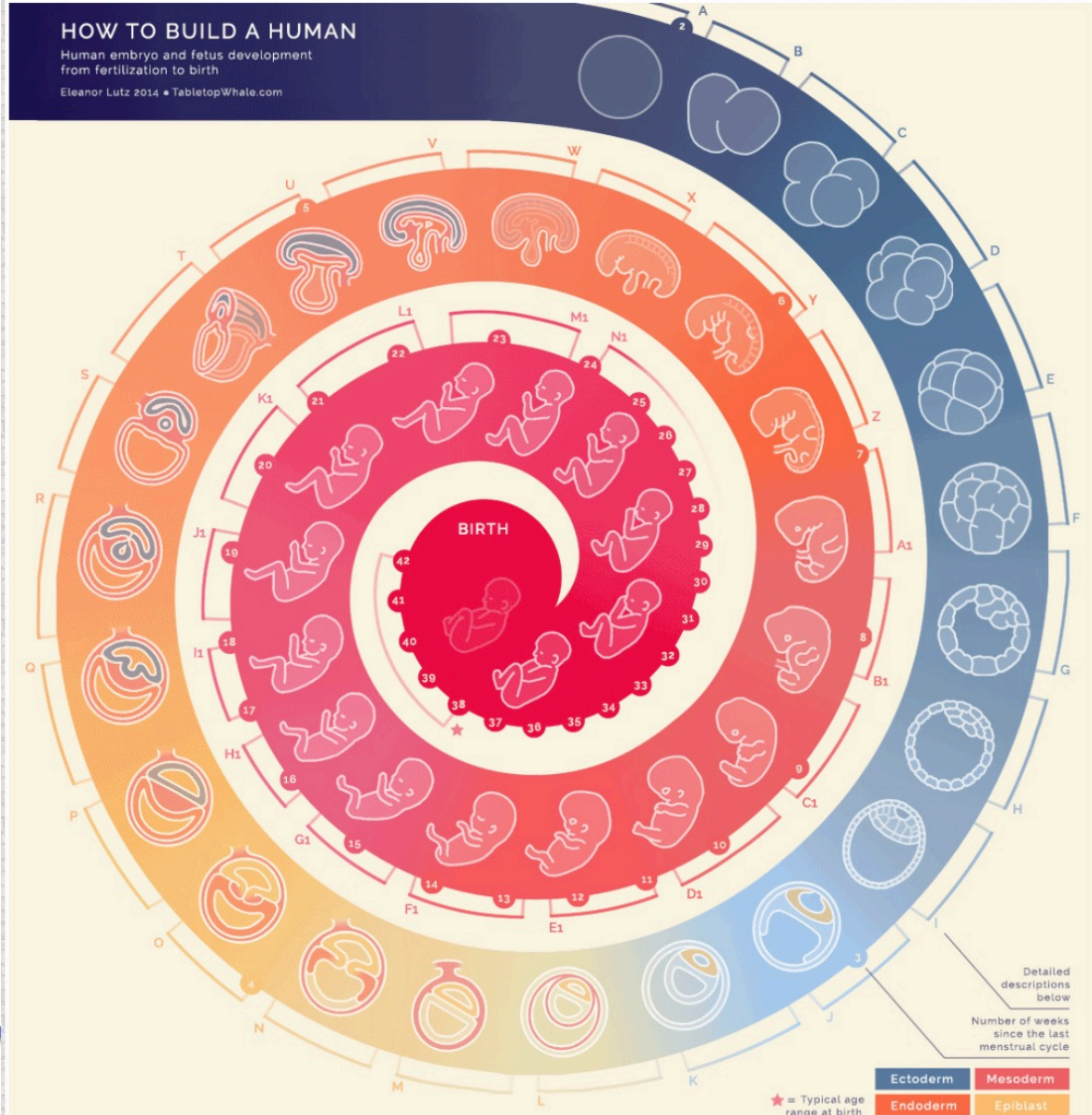
- impacts a particular application

- http://scimaps.org/



The tabletop whale

science illustration blog



Journalism

Alex Gonçalves: Visualization for the masses

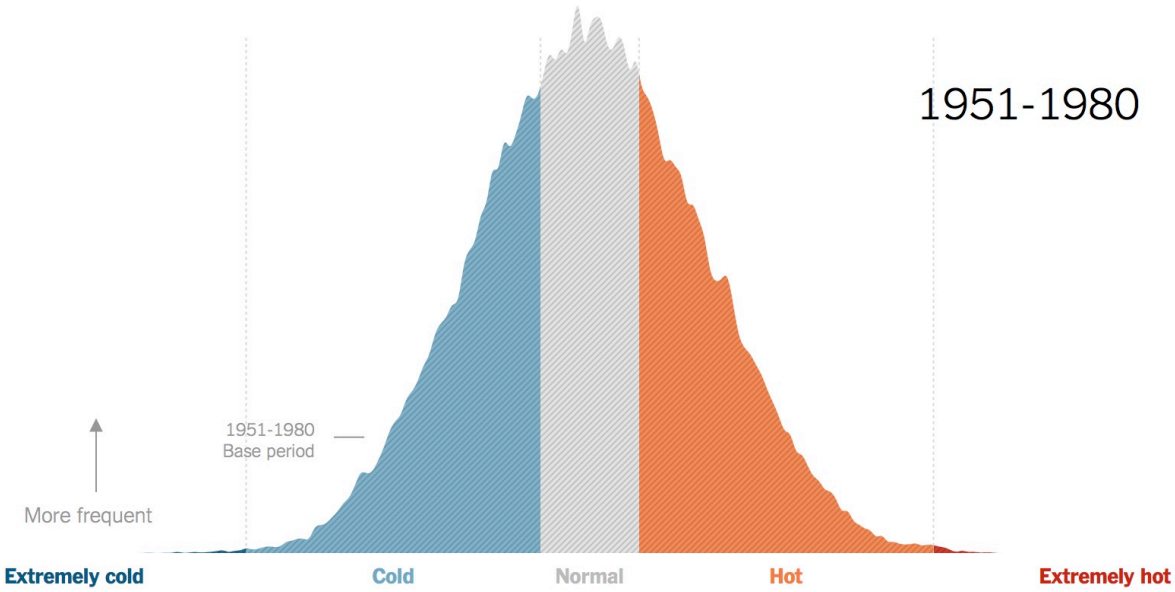
- to build communion
- for social change
- powerful stories
- “duty of beauty”

http://www.nytimes.com/newsgraphics/2014/02/14/fashion-week-editors-picks/

It's Not Your Imagination. Summers Are Getting Hotter.

By NADJA POPOVICH and ADAM PEARCE JULY 28, 2017

Summer temperatures
in the Northern Hemisphere



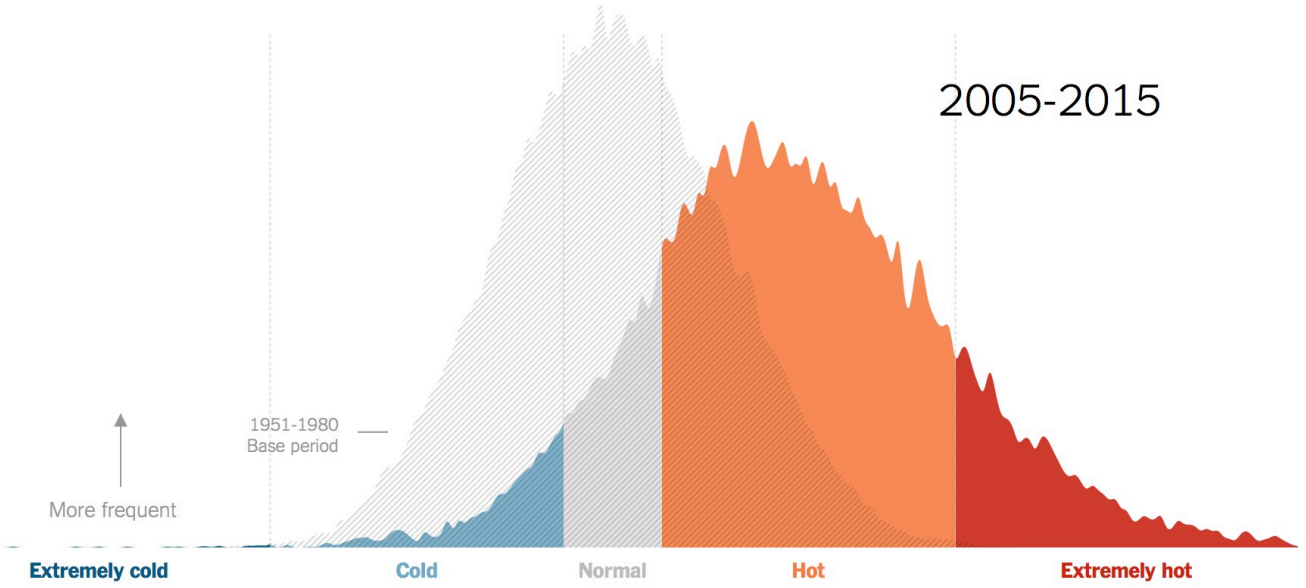
CLIMATE | It's Not Your Imagination. Summers Are Getting Hotter.

MORE 170

It's Not Your Imagination. Summers Are Getting Hotter.

By NADJA POPOVICH and ADAM PEARCE JULY 28, 2017

Summer temperatures
in the Northern Hemisphere



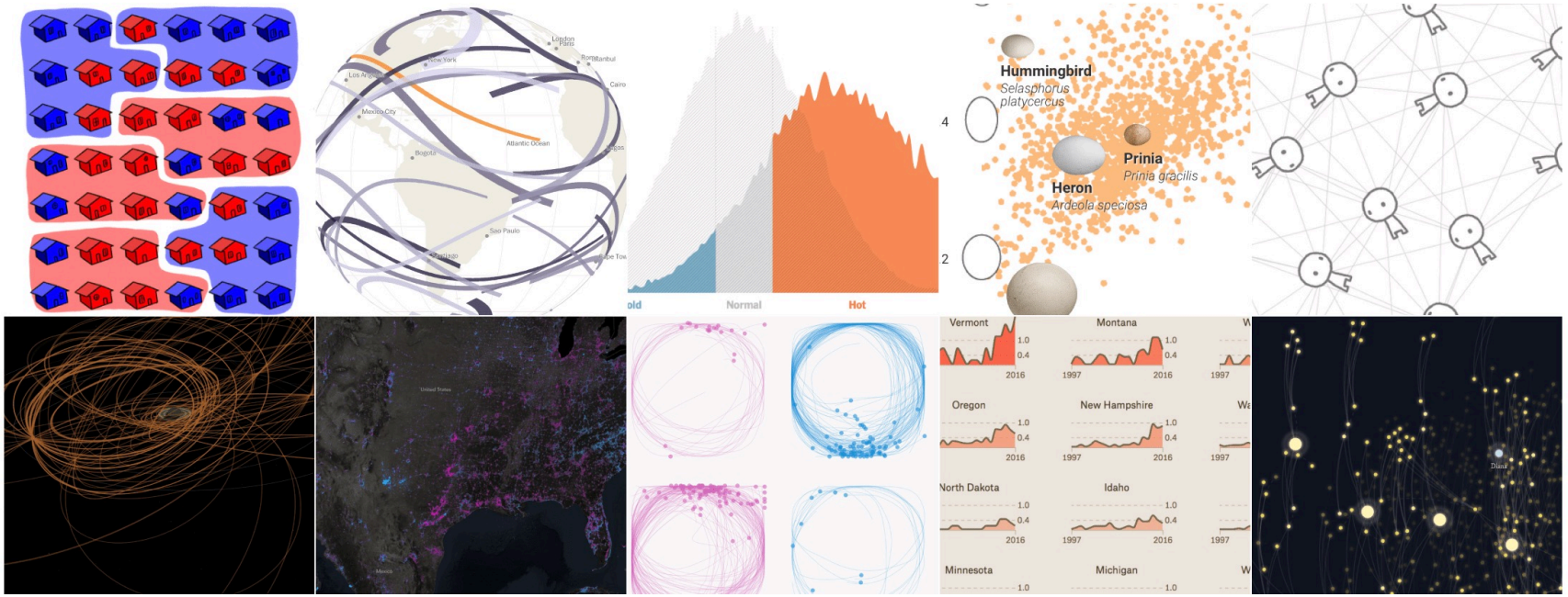
So much cool stuff!

FLOWINGDATA MEMBERSHIP COURSES TUTORIALS PROJECTS BECOME A MEMBER | LOG IN

VISUALIZATION / BEST-OF

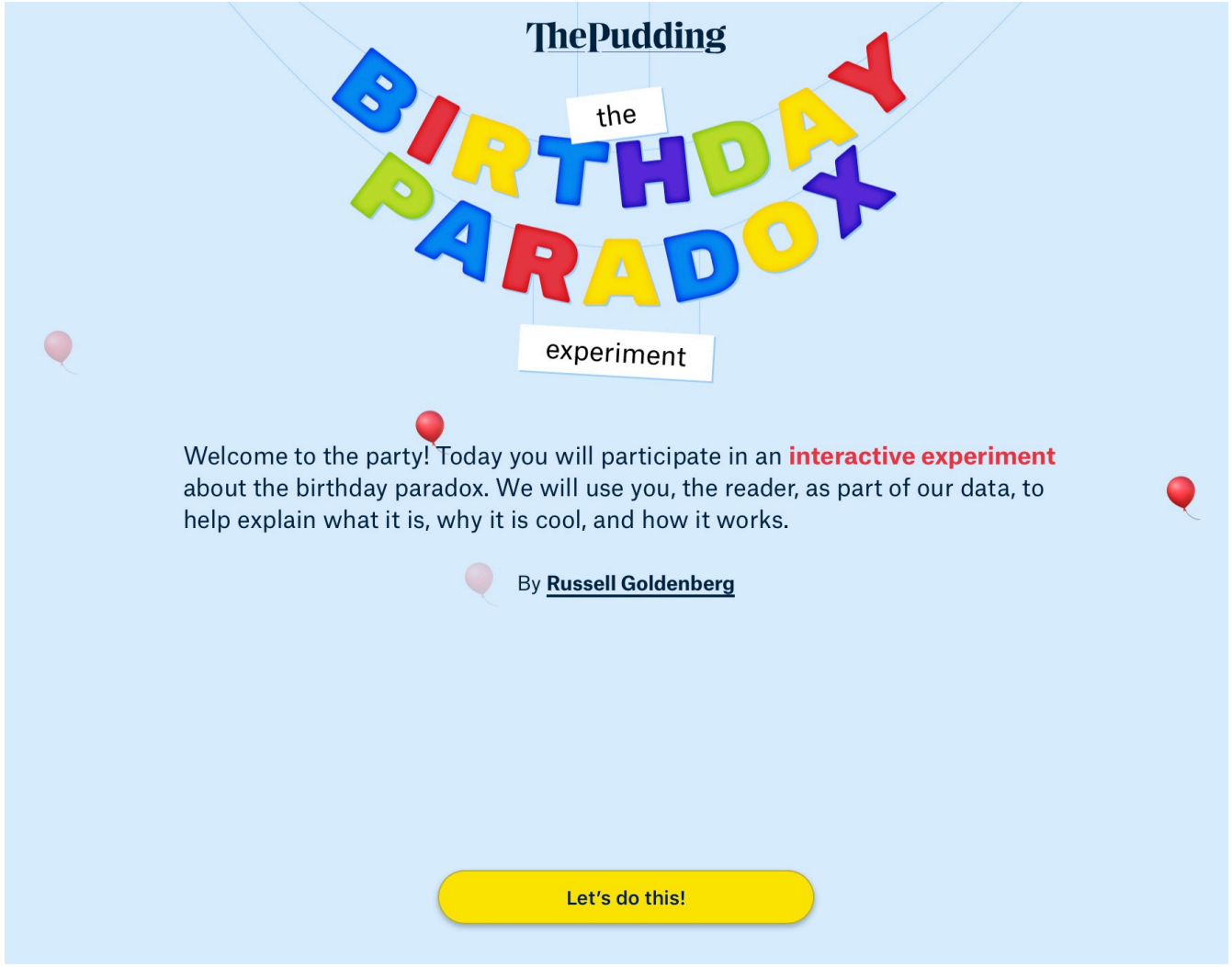
10 Best Data Visualization Projects of 2017

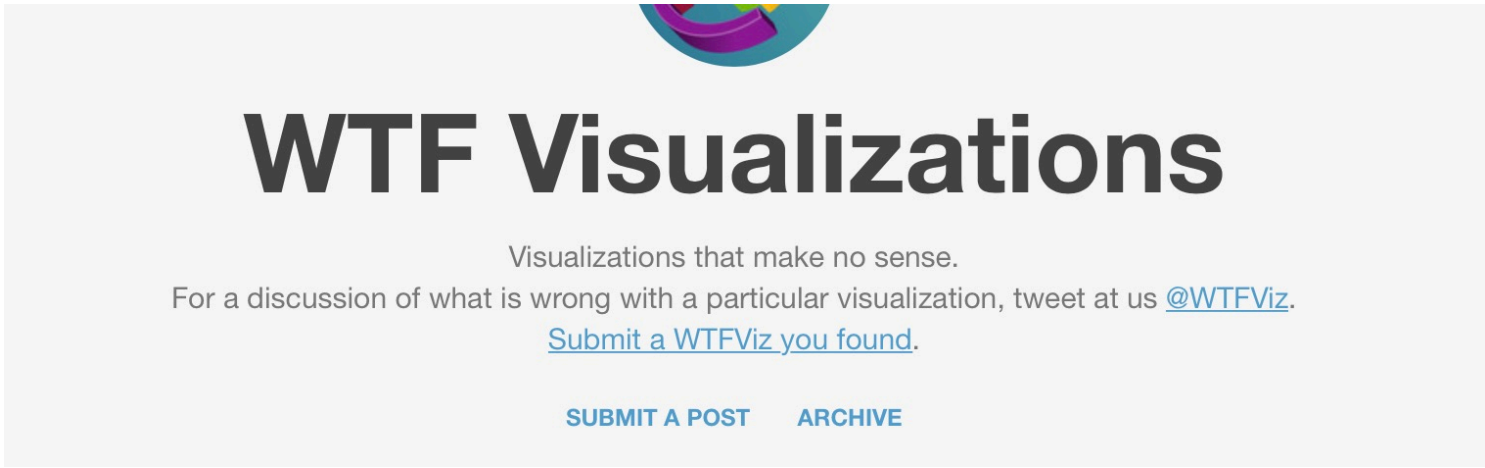
By Nathan Yau



So much cool stuff!

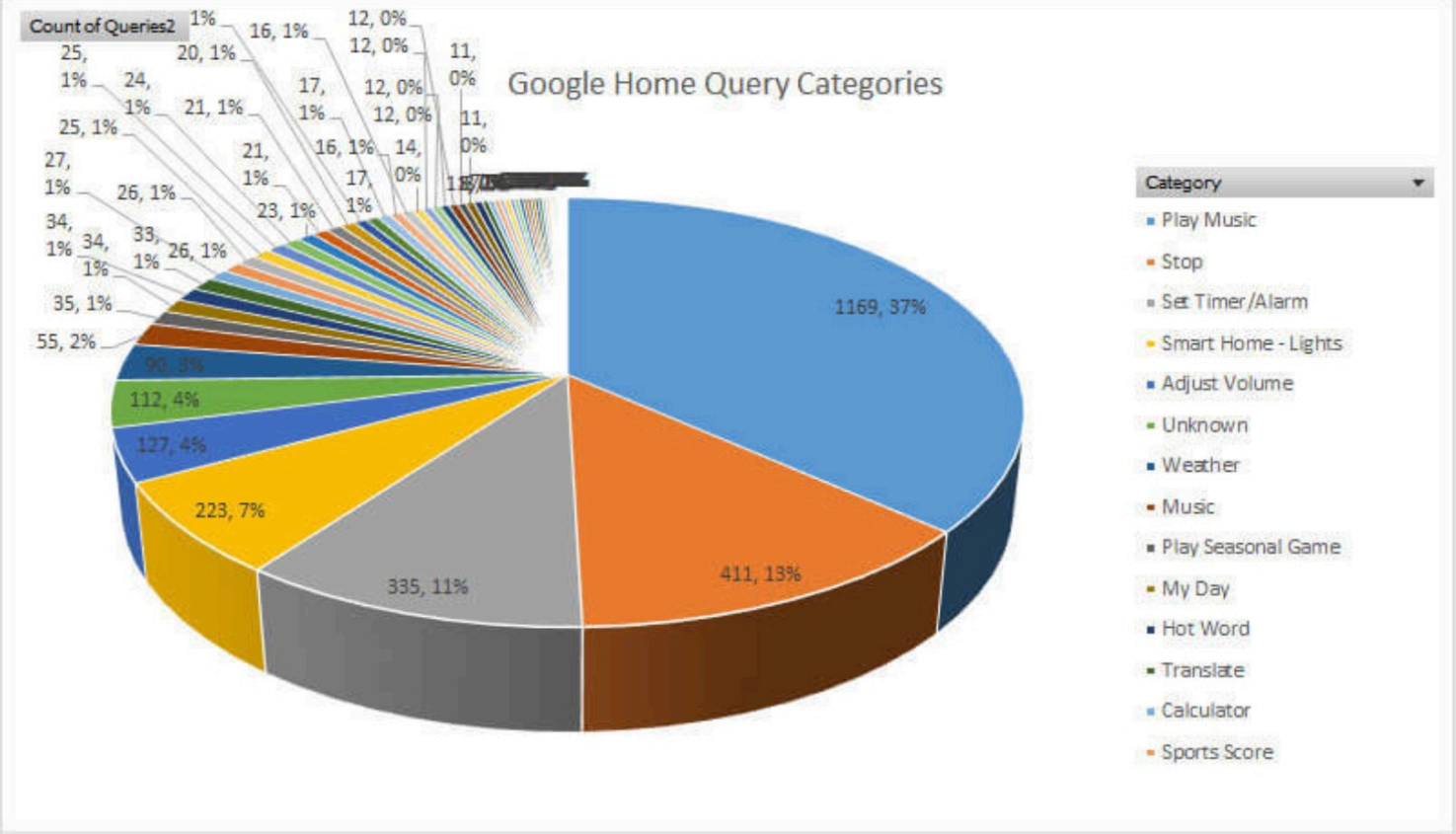
pudding.cool




WTF Visualizations
Visualizations that make no sense.
For a discussion of what is wrong with a particular visualization, tweet at us @WTFViz.
Submit a WTFViz you found.
SUBMIT A POST
ARCHIVE

WTF Visualizations

Visualizations that make no sense.



Just a sliver please.

The dreaded dashboard

Reese *Significance* April '18

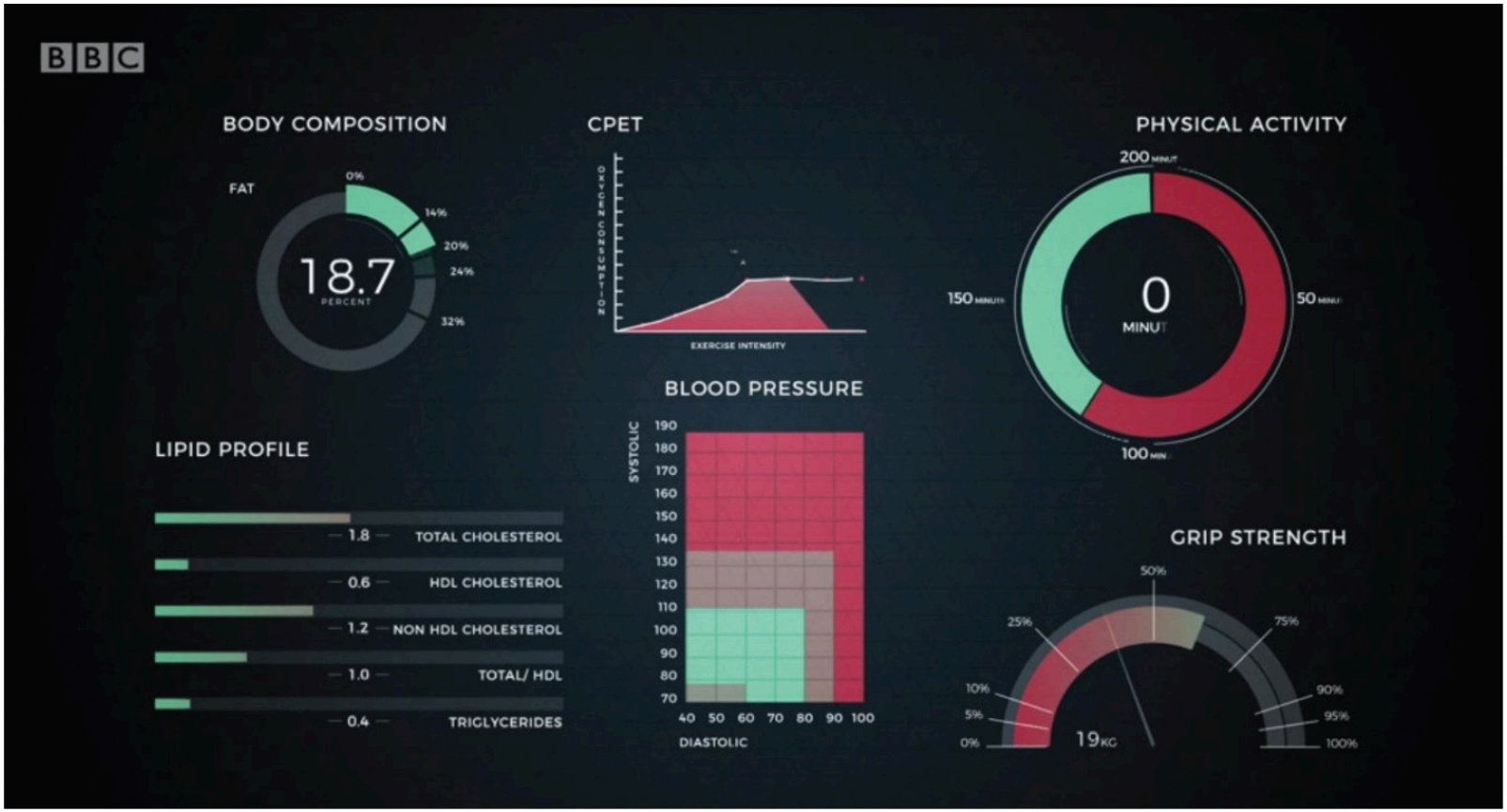


FIGURE 1 A selection of charts from *How to Stay Young*, showing some of the metrics used to calculate a person's body age, which can then be compared to their birth age.

The dreaded dashboard

Reese *Significance* April '18

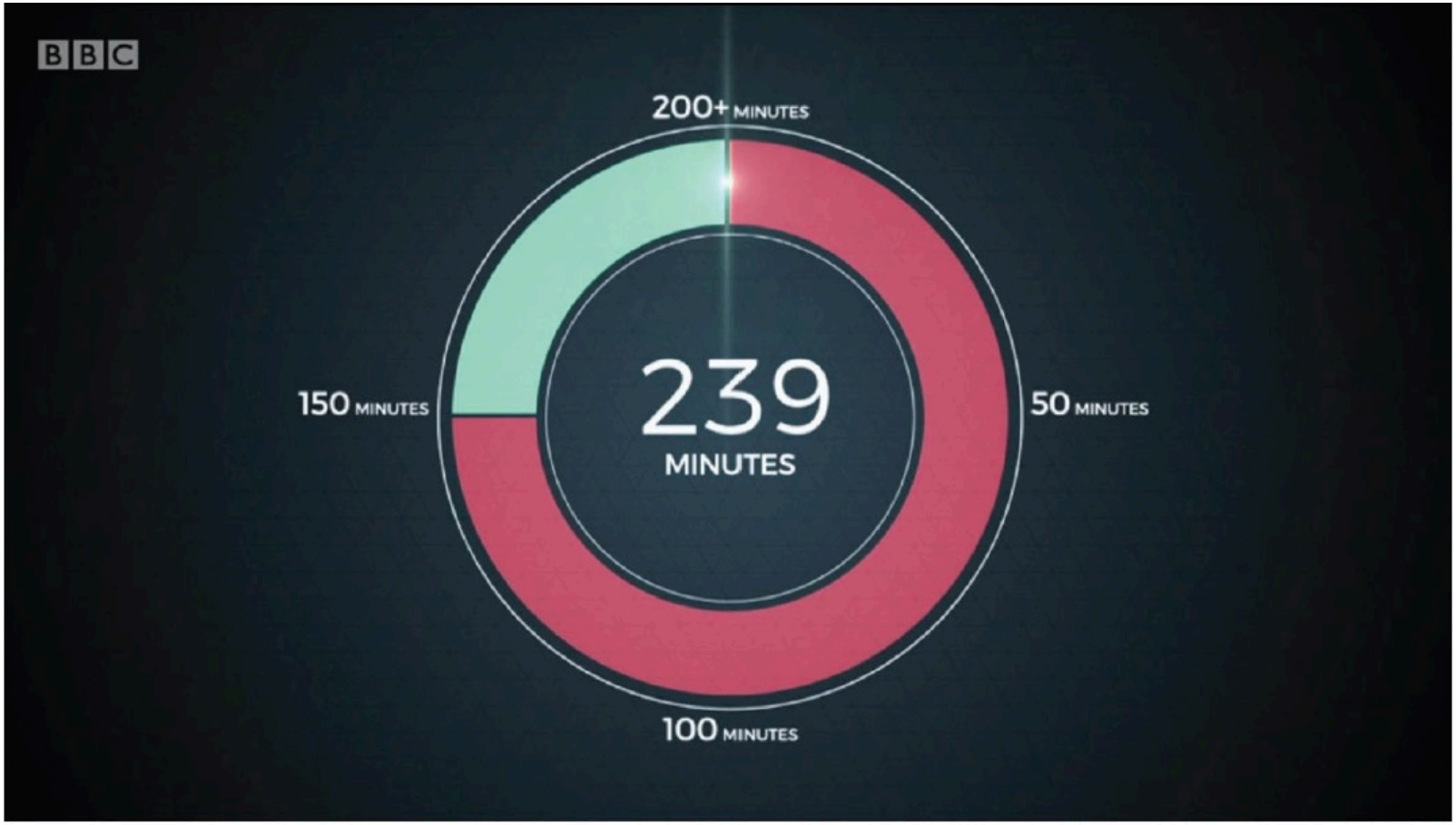


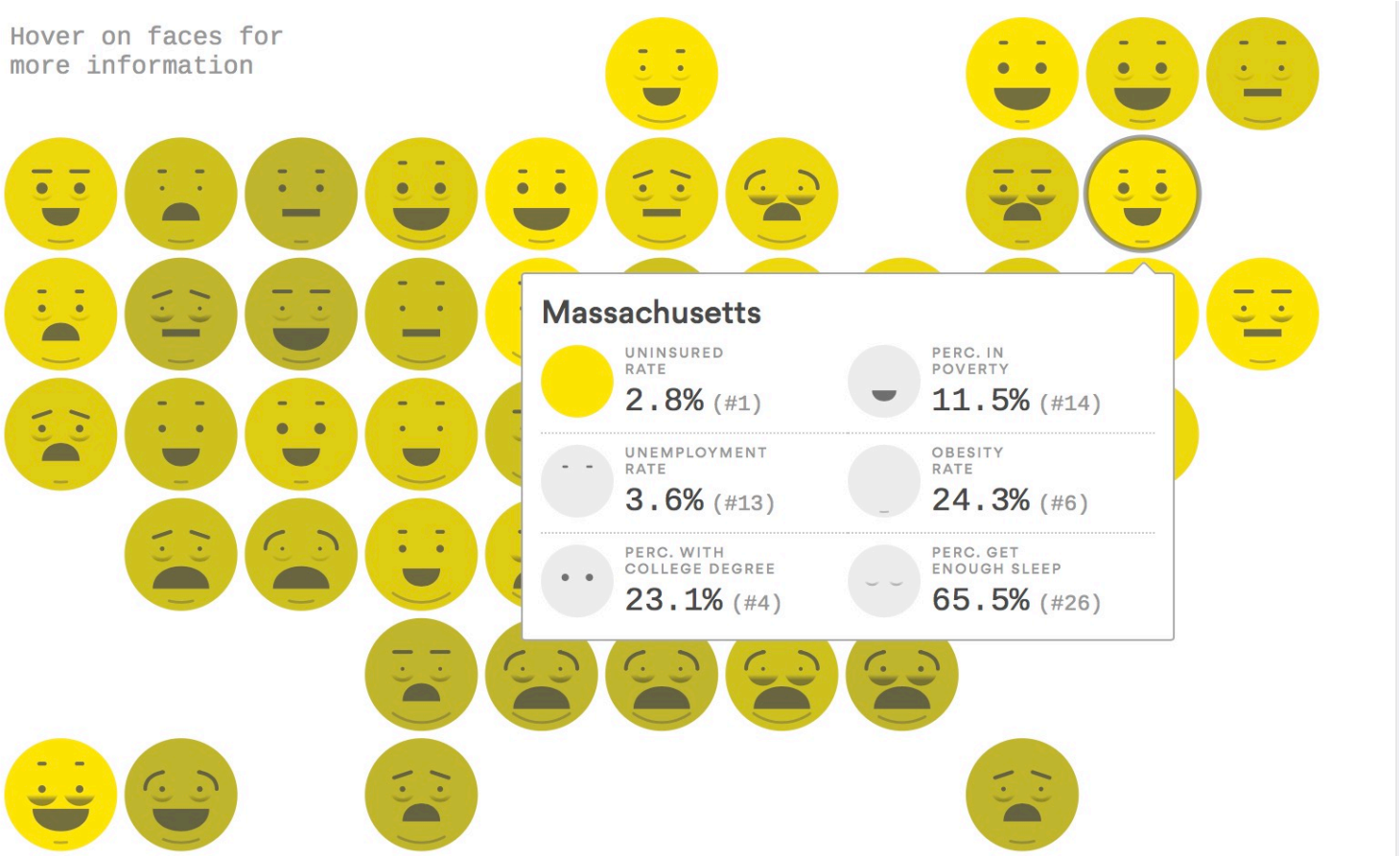
FIGURE 4 Minutes of exercise per week, with total in centre.

The Emoji States of America

axios.com

6 variables

The Emoji States of America



Thank you!

Thank you!

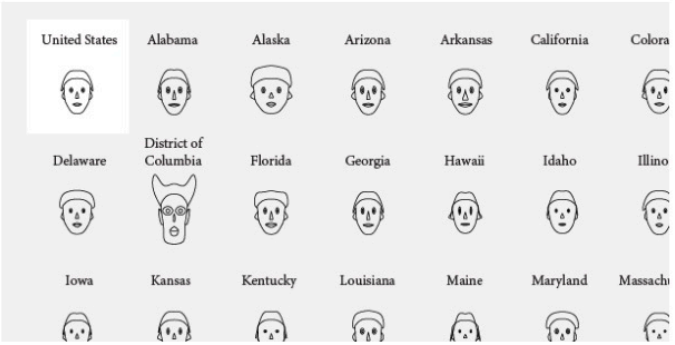
TUTORIALS / CHERNOFF FACES, MULTIVARIATE, R

How to visualize data with cartoonish faces ala Chernoff

By Nathan Yau

The goal of Chernoff faces is to show a bunch of variables at once via facial features like lips, eyes, and nose size. Most of the time there are better solutions, but the faces can be interesting to work with.

Demo | Download Source



Ross-Chernoff Glyphs Or: How Do We Kill Bad Ideas in Visualization?

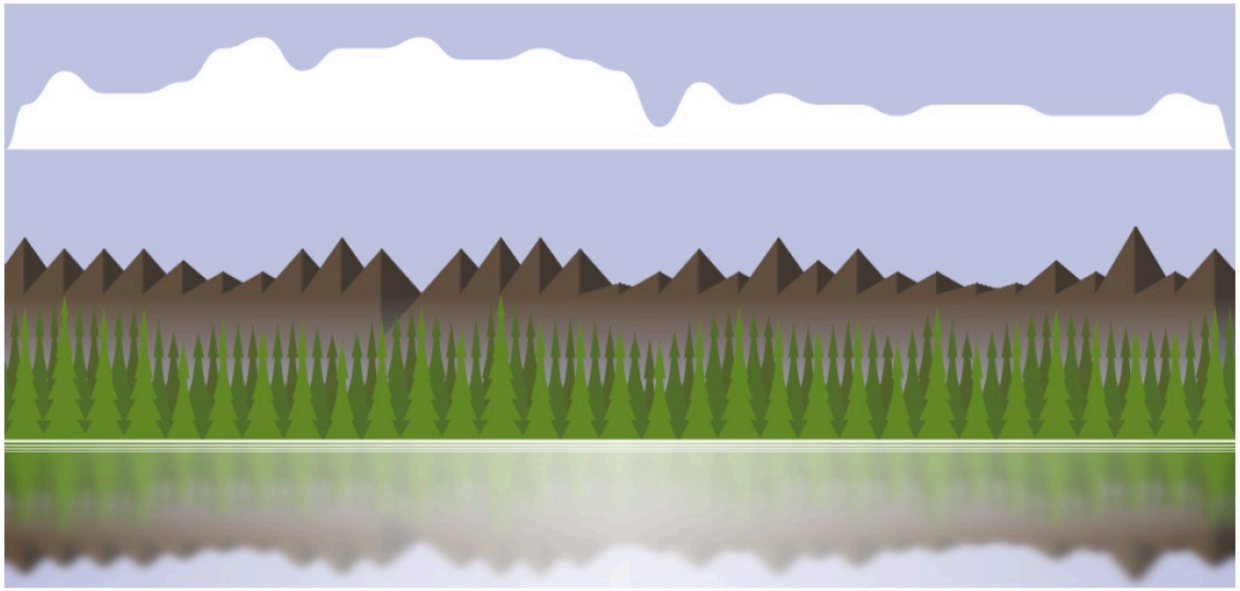


Figure 1: A meta-Ross-Chernoff visualization of Bob Ross's painting subjects, over time. Horizontal position is the season of "The Joy of Painting," the height of the treeline is the percentage of paintings depicting trees, the height of the mountains the percentage of paintings depicting mountains, and the height of the clouds is the percentage of paintings depicting clouds. From a dataset provided by FiveThirtyEight [12]