Ethical and Deceptive Visualization

Michael Correll
Tableau Research
Questions

What Makes A Visualization *Deceptive*?

What Makes a Visualization *Ethical*?
Questions

What Makes A Visualization Deceptive?

What Makes a Visualization Ethical?
HOW TO LIE WITH STATISTICS

DARRELL HUFF
Illustrated by Irving Geis
An Honest-to-Goodness Bestseller

HOW TO LIE WITH MAPS

MARK MONMONIER
Bad Visualizations

Incorrect Visualizations

Illegible Visualizations

Bullshit Visualizations

Deceptive Visualizations
Incorrect Visualization
Illegible Visualization
Bullshit Visualization

Lie:
“No, officer, I wasn’t speeding”
(you know the truth, but intentionally say something you know is untrue)

Bullshit:
“The party was lame anyways, it’s good I wasn’t invited”
(you don’t know or don’t care about the truth, but intentionally say something you hope is persuasive)
Number decoration

Moritz Stefaner

Truth & Beauty Operator

Editor of Visualizing, The Field

474 Following  51K Followers   ·   Twitter
The National Collegiate Health Assessment was taken by 1,000 UCSB students in Spring 2009. Participants were asked how frequently they used substances over the past 30 days. Numbers in white reflect actual student use, while red numbers indicate perceived substance use. The average age of participants was 20 years and approximately 99 percent were full-time students.
BY THE NUMBERS

The National Collegiate Health Assessment was taken by 1,000 UCSB students in Spring 2009. Participants were asked how frequently they used substances over the past 30 days. Numbers in white reflect actual student use, while red numbers indicate perceived substance use. The average age of participants was 20 years and approximately 99 percent were full-time students.

**OPIATES**
- 1-9 TIMES: 0.7%
- 10-29 TIMES: 29.1%
- DAILY: 2.4%
- Overall: 0.4%

**ALCOHOL**
- 1-9 TIMES: 56.9%
- 10-29 TIMES: 21.1%
- DAILY: 1.4%
- Overall: 11.2%

**COCAINE**
- 1-9 TIMES: 0.3%
- 10-29 TIMES: 0.2%
- DAILY: 0.0%
- Overall: 1.2%
BY THE NUMBERS

The National Collegiate Health Assessment was taken by 1,000 UCSB students in Spring 2009. Participants were asked how frequently they used substances over the past 30 days. Numbers in white reflect actual student use, while red numbers indicate perceived substance use. The average age of participants was 20 years, and approximately 99 percent were full-time students.

<table>
<thead>
<tr>
<th>Substance</th>
<th>1-9 Times</th>
<th>10-29 Times</th>
<th>Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opiates</td>
<td>0.7%</td>
<td>0%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Alcohol</td>
<td>56.9%</td>
<td>21.1%</td>
<td>11.2%</td>
</tr>
<tr>
<td>Cocaine</td>
<td>48.4%</td>
<td>0.2%</td>
<td>0.3%</td>
</tr>
</tbody>
</table>
UNEMPLOYMENT RATE
UNDER PRESIDENT OBAMA

JAN  FEB  MAR  APR  MAY  JUN  JUL  AUG  SEP  OCT  NOV
9.0%  8.9%  9.0%  9.0%  9.1%  9.2%  9.1%  9.1%  9.1%  9.1%  9.0%  8.6%
Deceptive Visualization
DARK PATTERNS

(They’re Everywhere)
Dark Patterns
Dark Patterns
Zone Kaiwei Ni factory
Sponsored
80% OFF
BLACK FRIDAY SALE
Dark Patterns in Visualization

Lie Factors

Scale Manipulation

Metric Manipulation
Tufte’s Lie Factor

Lie Factor =
The size of the effect shown in the graphic
The size of the effect in the data

LF ≅ 1 Good!
LF >> 1 or LF << 1 Bad!
Tufte’s Lie Factor

Lie Factor =
The size of the effect shown in the graphic
The size of the effect in the data

LF ≅ 1 Good!
LF >> 1 or LF << 1 Bad!

LF = \[\frac{(5.3 - 0.6)/0.6}{(27.5-18)/18}\]

LF = 14.8!
Tufte’s Lie Factor

\[
\text{Lie Factor} = \frac{\text{The size of the effect shown in the graphic}}{\text{The size of the effect in the data}}
\]

\[\text{LF} \approx 1 \text{ Good!}\]
\[\text{LF} >> 1 \text{ or LF} << 1 \text{ Bad!}\]
Lie Factor

Distorting the apparent size of the effect in your data, often through choosing ambiguous or non-standard encodings.
Scale Manipulation
IF BUSH TAX CUTS EXPIRE

TOP TAX RATE

NOW  JAN. 1, 2013

35%  39.6%
IF BUSH TAX CUTS EXPIRE

TOP TAX RATE

NOW  JAN. 1, 2013

35%   39.6%
“The Only Climate Change Graph You’ll Ever Need”
“The Only Climate Change Graph You’ll Ever Need”
Scale Manipulation

Changing with the scales of your chart to minimize, magnify, or invert the change in the data.
Convention Manipulation
Convention Manipulation

Gun deaths in Florida

Number of murders committed using firearms

- 2005: Florida enacted its ‘Stand Your Ground’ law

Source: Florida Department of Law Enforcement
C. Chan 16/02/2014
Convention Manipulation

Gun deaths in Florida

Number of murders committed using firearms

Source: Florida Department of Law Enforcement
C. Chan 16/02/2014
Convention Manipulation

**Graph: Gun deaths in Florida**

- Number of murders committed using firearms
- **2005**: Florida enacted its ‘Stand Your Ground’ law

Source: Florida Department of Law Enforcement

C. Chan 16/02/2014
Convention Manipulation

Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.
Convention Manipulation

Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.

<table>
<thead>
<tr>
<th>County</th>
<th>Cobb</th>
<th>DeKalb</th>
<th>Fulton</th>
<th>Gwinnett</th>
<th>Hall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deaths</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalizations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Convention Manipulation

Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.

Candice Broce
@candicebroce

Replying to @petecorson and @AJCInteractives

The x axis was set up that way to show descending values to more easily demonstrate peak values and counties on those dates. Our mission failed. We apologize. It is fixed.

1:38 PM · May 11, 2020 · Twitter for iPhone

14 Retweets 50 Quote Tweets 28 Likes
Convention Manipulation

Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases represents the number of deaths and hospitalizations in each of those impacted counties.

Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

- Cobb
- DeKalb
- Fulton
- Gwinnett
- Hall

Redesign from Lucy D'Agostino McGowan
https://livefreeordichotomize.com/2020/05/17/graph-detective/
Convention Manipulation

2016 Senate Seats

- R
- D
- I

47
Convention Manipulation

2016 Senate Seats

- R
- D
- I
Wrap Up

The designer of a chart can control the message of the chart.

There are patterns that can often result in misleading charts.

But critical thinking is more important than hard rules in design.
Questions

What Makes A Visualization *Deceptive*?

What Makes a Visualization *Ethical*?
Accounts of Ethics

Consequentialist Ethics
Deontological Ethics
Virtue Ethics
Role Ethics
... lots more
Consequentialism

A action is ethical or unethical based on its consequences. For instance, in “utilitarianism,” an action’s ethics is based on utility.
The Value of Visualization

- \(C_v(S_0)\): Initial development costs. The visualization method has to be developed and implemented, possibly new hardware has to be acquired.
- \(C_u(S_0)\): Initial costs per user. The user has to spend time on selection and acquisition of V, understanding how to use it, and tailoring it to his particular needs.
- \(C_t(S_0)\): Initial costs per session. Data have to be converted, and an initial specification of the visualization has to be made.
- \(C_e\): Perception and exploration costs. The user has to spend time to watch the visualization and understand it, as well as in modification and tuning of the specification, thereby exploring the data set.

The total costs are now given by

\[
C = C_t + nC_u + nmC_v + nmkC_e.
\]

The return on these investments consists of the value \(W(\Delta K)\) of the acquired knowledge \(\Delta K = K(T) - K(0)\) per session, multiplied by the total number of sessions:

\[
G = nmW(\Delta K)
\]

and hence for the total profit \(F = G - C\) we find

\[
F = nm(W(\Delta K) - C_t - kC_e) - C_t - nC_u.
\]

*Fig. 1: “A picture is worth a thousand words”*
The Value of Visualization

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Player In Year</td>
<td>NBA Stats</td>
<td>Assist To Turnover Percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Charles Barkley</td>
<td>1988</td>
<td>1.27592756</td>
<td>0.517942702</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Earnest Johnson</td>
<td>1988</td>
<td>1.26866867</td>
<td>0.50923483</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Michael Jordan</td>
<td>1988</td>
<td>2.24237931</td>
<td>0.58161516</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Karl Malone</td>
<td>1988</td>
<td>0.76842105</td>
<td>0.51802239</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Hakeem Olajuwon</td>
<td>1988</td>
<td>0.54182183</td>
<td>0.50771208</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Charles Barkley</td>
<td>1989</td>
<td>1.26537448</td>
<td>0.59832008</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Patrick Ewing</td>
<td>1990</td>
<td>0.65467626</td>
<td>0.51105158</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Earnest Johnson</td>
<td>1990</td>
<td>1.38940821</td>
<td>0.47978991</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Michael Jordan</td>
<td>1990</td>
<td>2.10223459</td>
<td>0.52467560</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Karl Malone</td>
<td>1990</td>
<td>0.74542255</td>
<td>0.56177013</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Charles Barkley</td>
<td>1990</td>
<td>1.35220805</td>
<td>0.56723119</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Earnest Johnson</td>
<td>1990</td>
<td>3.14986159</td>
<td>0.47749902</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Michael Jordan</td>
<td>1990</td>
<td>2.24275426</td>
<td>0.53820216</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Karl Malone</td>
<td>1990</td>
<td>1.06057389</td>
<td>0.52810129</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>David Robinson</td>
<td>1990</td>
<td>0.77050275</td>
<td>0.51509557</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Clyde Drexler</td>
<td>1991</td>
<td>2.13333333</td>
<td>0.4701897</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Michael Jordan</td>
<td>1991</td>
<td>2.445</td>
<td>0.3870347</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Karl Malone</td>
<td>1991</td>
<td>0.97177419</td>
<td>0.52658522</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Christopher Mullin</td>
<td>1991</td>
<td>1.43384538</td>
<td>0.53708999</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>David Robinson</td>
<td>1991</td>
<td>1.91456055</td>
<td>0.51120489</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Charles Barkley</td>
<td>1992</td>
<td>1.65226052</td>
<td>0.52034884</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Michael Jordan</td>
<td>1992</td>
<td>2.06723851</td>
<td>0.49472611</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Karl Malone</td>
<td>1992</td>
<td>1.28333333</td>
<td>0.53233255</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Hakeem Olajuwon</td>
<td>1992</td>
<td>1.11068702</td>
<td>0.53908511</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>William Price</td>
<td>1992</td>
<td>3.07142857</td>
<td>0.46773882</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Karl Malone</td>
<td>1993</td>
<td>1.40729046</td>
<td>0.49472216</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Hakeem Olajuwon</td>
<td>1993</td>
<td>1.83805257</td>
<td>0.51774408</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Scottie Pippen</td>
<td>1993</td>
<td>1.73706887</td>
<td>0.4906193</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Latrell Sprewell</td>
<td>1993</td>
<td>1.70539892</td>
<td>0.43260409</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>John Starks</td>
<td>1993</td>
<td>3.87593880</td>
<td>0.51278477</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Antawnm Hardaway</td>
<td>1994</td>
<td>2.13568892</td>
<td>0.52259196</td>
<td></td>
</tr>
</tbody>
</table>
Problems with consequentialism

What do we measure?

Who is doing the measuring?

When are we done measuring?
Deontological Ethics

An action is ethical or unethical based on whether or not it follows the rules.
Visualization “Rules”

Tufte’s Lie Factor

LF = size of effect shown in graphic / size of effect in data

“Dishonest” if LF!=1
Visualization “Rules”

Tufte’s Lie Factor
LF = size of effect shown in graphic / size of effect in data

LF = 2.8!
Visualization “White Lies”
Problems with deontology

Where do the rules come from?

Can the rules change?

What about edge cases?
Problems with deontology

Where do the rules come from?

Can the rules change?

What about edge cases?
Problems with deontology

Where do the rules come from?

Can the rules change?

What about edge cases?
Problems with deontology

Where do the rules come from?

Can the rules change?

What about edge cases?
Virtue Ethics

Your actions are assessed by the extent to which they cultivate your virtues.
Critical InfoVis: Exploring the Politics of Visualization

Disclosure
Plurality
Contingency
Empowerment

-Marian Dörk, Patrick Feng, Chris Collins, Sheelagh Carpendale
Problems with virtue ethics

Where do the virtues come from?

What do I do about conflicts in virtues?

What is never okay to do?
Role Ethics

Your actions are assessed by the extent to which you fulfill the obligations of your role.
Professional Ethics

ACM CODE OF ETHICS

Computing professionals' actions CHANGE THE WORLD. They should reflect upon the wider impact of their work, consistently supporting the public good.

GENERAL ETHICAL PRINCIPLES.
Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing. Avoid harm. Be honest and trustworthy. Be fair and take action not to discriminate. Respect the work required to produce new ideas, inventions, and creative works, and computing artifacts. Respect privacy. Honor confidentiality.

PROFESSIONAL RESPONSIBILITIES.
Strive to achieve high quality in both the processes and products of professional work. Maintain high standards of professional competence, conduct, and ethical practice. Know and respect existing rules pertaining to professional work. Accept and provide appropriate professional reviews. Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks. Perform work only in areas of competence. Foster public awareness and understanding of computing, related technologies, and their consequences. Access computing and communication resources only when authorized or when compelled by the public good. Design and implement systems that are robustly and usability secure.

PROFESSIONAL LEADERSHIP PRINCIPLES.
Ensure that the public good is the central concern during all professional computing work. Articulate, encourage acceptance of, and evaluate fulfillment of social responsibilities by members of the organization or group. Manage personnel and resources to enhance the quality of working life. Articulate, apply, and support policies and processes that reflect the principles of the Code. Create opportunities for members of the organization or group to grow as professionals. Use care when modifying or retiring systems. Recognize and take special care of systems that become integrated in the infrastructure of society.

https://www.acm.org/code-of-ethics
Visualization Hippocratic Oath?

“I shall not use visualization to intentionally hide or confuse the truth which it is intended to portray. I will respect the great power visualization has in garnering wisdom and misleading the uninformed. I accept this responsibility willfully and without reservation, and promise to defend this oath against all enemies, both domestic and foreign.”

-Jason Moore
Problems with role-based ethics

Who gives you your roles?

What happens if you have multiple roles?

How big do our roles get?
Accounts of Ethics

Consequentialist Ethics
Deontological Ethics
Virtue Ethics
Role Based Ethics
... lots more
Berkeley/Haas Ethics Questions

Would I be happy for this decision to be on the public record?

What would happen if everybody did this?

How would I like it if someone did this to me?

Will the proposed course of action bring about a good result?

What will the proposed course of action do to my character or the character of my organization?

Is the proposed course of action consistent with my espoused values and principles?
What’s a “Good” Visualization?

One that has the most *benefits* for the least *cost*

One that follows the *rules of good design*

One that *cultivates* the right *virtues*

One that *fulfills our duties* as visualization designers
Manifestos

I SWEAR TO SUBMIT TO THE FOLLOWING SET OF RULES

1. Location
   a. The sound must never be music.
   b. The cameraman must not speak to the actors.
   c. The film must be shot on location.
2. Hand-held camera
   a. Hand-held cameras are necessary.
   b. The sound must never be music.
3. Color
   a. Special lighting is not acceptable.
   b. The sound must be listened to as music.
4. Forbidden
   a. No dialogue allowed.
   b. The sound must be listened to as music.
5. Genre movies
   a. No sound effects allowed.
   b. The sound must be listened to as music.

I AM HEREBY

I SWEAR TO ABIDE BY THE FOLLOWING RULES

1. I swear to submit to the following set of rules.
2. I swear to abide by the sound of my own voice.
3. I swear to submit to the sound of my own voice.
4. I swear to abide by the sound of my own voice.

I AM HEREBY

I SWEAR TO ABIDE BY THE FOLLOWING RULES

1. I swear to submit to the following set of rules.
2. I swear to abide by the sound of my own voice.
3. I swear to submit to the sound of my own voice.
4. I swear to abide by the sound of my own voice.

Courage, audacity, and revolt will be essential elements of our poetry.

We will destroy the museums, libraries, academies of every kind, will fight moralism, feminism, every opportunistic or utilitarian cowardice.
Data Humanism

“We are ready to question the impersonality of a merely technical approach to data and to begin designing ways to connect numbers to what they really stand for: knowledge, behaviors, people.”

-Giorgia Lupi
Data Feminism

“We call attention to the people and their bodies who are typically included in the data collection process, as well as to the people and their bodies who are typically left out.”

-Lauren Klein and Catherine D’Ignazio
Data for Black Lives

“White supremacy is an algorithm, in and of itself... much of the work for us as data practitioners is to reverse engineer that algorithm.”

-Ejike Obineme
Human-Computer Insurrection: Notes on an Anarchist HCI

“We must work to ensure our technologies actively contribute to (rather than detract from) human autonomy and dignity.”

-Os Keyes, Josephine Hoy, and Margaret Drouhard
Questions to Ask

Can I identify ethical and unethical data visualizations?

Can I critique the moral character of data visualizations?

Can I impact (or stop) projects I think are unethical?
Thanks!

Michael Correll

correll.io
Why is our conception of “bad” visualization so narrow?
CHANGING FACE OF AMERICA
Percent of total U.S. population by race and ethnicity, 1960-2060

1960
- Other: 85%
- Asian: 10%
- Hispanic: 5%
- Black: 1%
- White: 8%

2010
- Other: 85%
- Asian: 16%
- Hispanic: 12%
- Black: 1%
- White: 64%

2060
- Other: 85%
- Asian: 31%
- Hispanic: 13%
- Black: 1%
- White: 43%

Source: Pew Research Center

NIGHTLY NEWS
The Ethernet Delusion

“The purpose of visualization is to maximize throughput and minimize noise as you transport a dataset to a person’s brain”
Problems with the Ethernet Delusions

Ignores the source and rhetorical objective of visualization

Assumes data are objective truths about the world

Pre-supposes that you *must* create a visualization
<table>
<thead>
<tr>
<th>Call Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crimes Against Persons</td>
<td>21</td>
</tr>
<tr>
<td>Assault</td>
<td>8</td>
</tr>
<tr>
<td>Robbery</td>
<td>5</td>
</tr>
<tr>
<td>Threats</td>
<td>8</td>
</tr>
<tr>
<td>Assisting the Public</td>
<td>18</td>
</tr>
<tr>
<td>Drugs and Vice</td>
<td>17</td>
</tr>
<tr>
<td>Property Crime/Theft</td>
<td>75</td>
</tr>
</tbody>
</table>
White Collar Crime Risk Zones uses machine learning to predict where financial crimes are mostly likely to occur across the US. To learn about our methodology, read our white paper.

By Brian Clifton, Sam Lavigne and Francis Tseng for The New Inquiry Magazine, Vol. 59: ABOLISH.