Telling stories with data

Maish Nichani
@maish
PM mentions data visualization at National Day Rally 2016
Ignaz Semmelweis

1818-1865

https://en.wikipedia.org/wiki/Ignaz_Semmelweis
Statue of Semmelweis in front of Szent Rókus Hospital, Budapest, Hungary.

https://en.wikipedia.org/wiki/Ignaz_Semmelweis
Napoleon’s ill-fated march to Moscow
3 elements of data storytelling

- Data
- Narrative
- Visuals

Story of change
Story of change
“The purpose of visualisation is insight, not pictures.”

- Ben Shneiderman
  Pioneer in human-computer interaction
- Who is your audience
- What are you trying to communicate
- Why should they care
Where’s the story?
Where’s the story?
Exploratory analysis ➔ Interesting find ➔ Explanatory analysis
1. Know your audience
2. Pick one or two points that matter
3. Pick a perspective or angle that will create interest or spur curiosity
Which country won the most medals?

<table>
<thead>
<tr>
<th>Country</th>
<th>Gold</th>
<th>Silver</th>
<th>Bronze</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>46</td>
<td>37</td>
<td>38</td>
<td>121</td>
</tr>
<tr>
<td>Great Britain</td>
<td>27</td>
<td>23</td>
<td>17</td>
<td>67</td>
</tr>
<tr>
<td>China</td>
<td>26</td>
<td>18</td>
<td>26</td>
<td>70</td>
</tr>
<tr>
<td>Russia</td>
<td>19</td>
<td>18</td>
<td>19</td>
<td>56</td>
</tr>
<tr>
<td>Germany</td>
<td>17</td>
<td>10</td>
<td>15</td>
<td>42</td>
</tr>
<tr>
<td>Singapore</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Data shows that the usual suspects will be on the top of the table: United States, China, Britain and Russia. Population and money are two factors that put those countries on top of all others. But, what would happen if we measure all nations equally?

<table>
<thead>
<tr>
<th>BY GROSS DOMESTIC PRODUCT</th>
<th>BY GDP PER CAPITA</th>
<th>BY POPULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Medals per $100 million of GDP</strong></td>
<td><strong>Medals per $1,000 of GDP per capita</strong></td>
<td><strong>Medals per 1 million inhabitants</strong></td>
</tr>
<tr>
<td>Granada (1)</td>
<td>China (70)</td>
<td>Grenada (1)</td>
</tr>
<tr>
<td>Jamaica (11)</td>
<td>Ethiopia (8)</td>
<td>Bahamas (2)</td>
</tr>
<tr>
<td>Georgia (7)</td>
<td>Kenya (13)</td>
<td>New Zealand (18)</td>
</tr>
<tr>
<td>Armenia (4)</td>
<td>North Korea (7)</td>
<td>Jamaica (11)</td>
</tr>
<tr>
<td>Burundi (1)</td>
<td>Russia (56)</td>
<td>Denmark (15)</td>
</tr>
<tr>
<td>Azerbaijan (18)</td>
<td>United States (121)</td>
<td>Croatia (10)</td>
</tr>
<tr>
<td>North Korea (7)</td>
<td>Uzbekistan (13)</td>
<td>Slovenia (4)</td>
</tr>
<tr>
<td>Bahamas (2)</td>
<td>Britain (67)</td>
<td>Azerbaijan (18)</td>
</tr>
<tr>
<td>Serbia (8)</td>
<td>Ukraine (11)</td>
<td>Hungary (15)</td>
</tr>
<tr>
<td>Kenya (13)</td>
<td>Jamaica (11)</td>
<td>Bahrain (2)</td>
</tr>
</tbody>
</table>

https://www.washingtonpost.com/graphics/sports/olympics/rio-2016-olympic-medal-count/
People are asking why it is so difficult to locate the backbox of the crashed Malaysian Airlines flight MH370?

You have one data point. The black box could be as deep as 15,000 feet.

How would you explain this story?
The depth of the problem

After an Australian vessel, Ocean Shield, again detected deep-sea signals consistent with those from an airplane’s black box, the official leading a multination search expressed hope Wednesday that crews will begin to find wreckage of a missing Malaysian airliner “within a matter of days.”

“I believe we’re searching in the right area,” Retired Air Chief Marshal Angus Houston said.

All commercial transport aircraft are fitted with underwater locator beacons to assist in the relocation of black box flight data recorders and cockpit voice recorders. These beacons are free-running pingers that transmit signals at an acoustic frequency of 37.5 kilohertz and have an expected battery life of 30 days. The scale of the challenge in locating the black boxes is immense.
1. Look for interesting perspectives
2. Gather the right core data
3. Mashup with relevant data
The Functional Art by Alberto Cairo
Welcome to the Graph Design I.Q. Test.

This test will ask you 10 questions to determine how well you understand the principles of good table and graph design. Good luck!
Kurt Vonnegut on the **Shapes of Stories**
A statistical story conveys a message that tells readers what happened, who did it, when and where it happened, and hopefully, why and how it happened.

A statistical story can
- provide general awareness/perspective/context;
- inform debate on specific issues.

Fractions of a Second: An Olympic Musical

At the Olympics, the blink of an eye can be all that separates the gold medalist from the 10th-place finisher. In some events, this is obvious. But in others, with athletes racing one by one, the closeness of the race is harder to perceive. Listen to the differences below.

Alpine skiing

The women's downhill course was extremely tiring, and, because it was more challenging than the men's course, it ended up separating the skiers by much larger margins. This pattern appears in the two speed events: the downhill and the super-G.
<table>
<thead>
<tr>
<th><strong>Headline</strong></th>
<th><strong>Make them compelling</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Readers are most likely to read the headline before deciding to read the full story. Therefore, it should capture their attention. The headline should be short and make people want to read on. It should say something about the findings presented in the article, not just the theme.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Lead paragraph</strong></th>
<th><strong>Unpack the story</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>It summarizes the story line concisely, clearly and simply. It should contain few numbers. In fact, try writing the first sentence of the lead using no figures at all.</td>
</tr>
</tbody>
</table>

Prices up in domestic and import markets

Gasoline prices hit 10-year low


Despite mounting financial challenges during the 1990s, young people from moderate and low-income families were no less likely to attend university in 2001 than they were in 1993, according to a new study.
The largest contributor to the monthly increase in the CPI was a 0.5% rise in the transportation index. Higher auto insurance premiums and air fares helped push up consumer prices this month.
Driving Shifts Into Reverse

ECONOMISTS have long studied the relationship between driving habits and gasoline prices. Low gas prices have long periods of profligate driving, and a quick jump in prices can cause many drivers to hoard gasoline for later use. Until recently, Americans have driven more each year than the previous one, with a few brief exceptions. In 1974, Americans of driving age drove about 4000 miles a year, on average. Fifty years later, the figure had climbed above 12000.

But the latest recession has caused some big changes. High unemployment meant that fewer people were driving to work, and a drop in consumer spending meant that less freight needed to be moved around the country. As gas prices soared in 2008, the number of miles driven — including commercial and personal — began to fall, and continued to drop after 2009 even as gasoline became cheaper.

"People were surprised by the very rapid rise in gas prices, and they changed their driving behavior," said Kenneth A. Small, a transportation economist at the University of California, Irvine. "But my suspicion is that it is temporary. As soon as unemployment gets back to pre-recession levels, we will see Americans doing a lot more driving again."

Source: Design Information Administration, Federal Highway Administration, InfoGraphics Institute
Let’s practice
What is the objective?
Who is the audience?
Why should they care?
How is the story composed?

Image from "Iraq's Bloody Toll", published in the South China Morning Post on Saturday, December 13, 2011
● What is the objective?
● Who is the audience?
● Why should they care?
● How is the story composed?

http://www.wsj.com/articles/SB10001424052748703338004575230041742556522