04 | from exploring to explaining: experiments with data-ink ratios; annotations; hierarchy; audiences and complexity; (re)design



Scott Spencer | Columbia University

course overview, learn to drive change using data visuals and narrative





general course deliverable timeline

Individual Work

For learning data visualization and written narrative techniques

Sept 30	Oct 14	Oct 28	Nov 18	Nov 18	Dec 11	
Homework 1 graphics	Homework 2 graphics	Homework 3 writing	Homework 4 graphics	Proposal	Interactive Communication	Multimodal commur
10%	10%	10%	10%	15%	20%	15%
				Participation 10%		

Group work

For building graphics and narrative into interactive communications







next deliverable, homework two

Individual Work

For learning data visualization and written narrative techniques

Sept 30 Oct 14		Oct 28	Nov 18	Nov 18	Dec 11	1		
Homework 1 graphics	Homework 2 graphics	Homework 3 writing	Homework 4 graphics	Proposal	Interactive Communication	Multimodal commun		
10%	10%	10%	10%	15%	20%	15%		
				Participation 10%	1			

For building graphics and narrative





ication	

homework two check-in | graphics practice with Citi Bike rebalancing study

homework two check-in, questions?



Saldarriaga, Juan Francisco. CitiBike Rebalancing Study. Spatial Information Design Lab, Columbia University, 2013. https:// c4sr.columbia.edu/projects/citibike-rebalancing-study.



© 2021 Scott Spencer / 🎧 https://ssp3nc3r.github.io

scott.spencer@columbia.edu

Doumont's three laws of communication applied to data encoding

Doumont applied to data encoding, Doumont's three laws of communication

Adapt to your audience

Maximize the signal-to-noise ratio

Use effective redundancy





Doumont applied to data encoding, "data ink" — Jacques Bertin's visual channels for encoding data



		Title of the Char can be Multiple I Subtitle in sentence	t Lines. e case.			
	axis.text.y -					
	axis.text.y -					
	➤ axis.text.y -					
	axis.text.y -					
	axis.text.y -					
		axis.text.x	axis.text.x	axis.text.x X	axis.text.x Caption text can be use or, say, describing the d	axis.text.x ed for notes ata source.



Doumont applied to data encoding, the data ink — example functions to draw encoded data in ggplot2

```
# load grammar of graphics
library(ggplot2)
p <-
  # functions for data ink
  ggplot(data = <data>,
          mapping = aes(<aesthetic> = <variable>,
                          <aesthetic> = <variable>,
                          \langle \ldots \rangle = \langle \ldots \rangle +
  geom_<type>(<...>) +
  scale_<mapping>_<type>(<...>) +
  coord_<type>(<...>) +
  facet_<type>(<...>) +
  \langle \ldots \rangle +
                                              element_blank()
  # functions for non-data ink
                                              element_line(<...> = <...>)
                                              element_rect(<...> = <...>)
  labs(<...>) +
  theme(<...> = <...>) +
                                              element_text(<...> = <...>)
  annotate(<...>) +
```

<...>



scott.spencer@columbia.edu



Doumont applied to data encoding, non-"data ink"





Doumont applied to data encoding, non-"data ink" — example functions to draw non-data ink in ggplot2

```
# load grammar of graphics
library(ggplot2)
p <-
  # functions for data ink
  ggplot(data = <data>,
          mapping = aes(<aesthetic> = <variable>,
                         <aesthetic> = <variable>,
                         \langle \ldots \rangle = \langle \ldots \rangle +
  geom_<type>(<...>) +
  scale_<mapping>_<type>(<...>) +
  coord_<type>(<...>) +
  facet_<type>(<...>) +
  <...>+
                                             element_blank()
  # functions for non-data ink
                                             element_line(<...> = <...>)
                                             element_rect(<...> = <...>)
  labs(<...>) +
  theme(<...> = <...>) +
                                             element_text(<...> = <...>)
  annotate(<...>) +
```

<...>



scott.spencer@columbia.edu



Doumont applied to data encoding, Tufte — data-ink maximization, within reason

data-ink data-ink ratio = total ink used to print the graphic

> proportion of a graphic's ink devoted to the =non-redundant display of data-information

1.0 – proportion of a graphic that can be = erased without loss of data-information

Doumont — "maximize the signal-to-noise ratio"

"The principle *helps conduct experiments* in graphical design."

— Tufte, Edward, *The Visual Display of Quantitative Information*

© 2021 Scott Spencer / 🌍 https://ssp3nc3r.github.io



scott.spencer@columbia.edu



Doumont applied to data encoding, example — removing gridlines increases the data-ink ratio



Doumont — "maximize the signal-to-noise ratio"



scott.spencer@columbia.edu



















15% -









© 2021 Scott Spencer / 💭 https://ssp3nc3r.github.io



Respense scott.spencer@columbia.edu









"In these revisions of the box plot, . . . the best overall arrangement naturally also rests on statistical and aesthetic criteria — in other words, the procedure is one of *reasonable* data-ink maximizing."

— Tufte, Edward, *The Visual Display of Quantitative Information*

© 2021 Scott Spencer / 🎧 https://ssp3nc3r.github.io



© 2021 Scott Spencer / 💭 https://ssp3nc3r.github.io



© 2021 Scott Spencer / 💭 https://ssp3nc3r.github.io











Doumont — "adapt to your audience"



One empirical experiment, asking participants to state data characteristics, given one of these randomized versions of the boxplots shown here, suggested that one of Tufte's erasures (option C) was most cognitively difficult for viewers to interpret.

Cues like labels and gridlines, together with some strictly superfluous embellishment of data points or other design elements, may sometimes be an aid rather than an impediment to interpretation.

Adapt to your audience.





Doumont applied to data encoding, data-ink maximization — one of many design considerations

"Maximizing data ink (within reason) is but a single dimension of a complex and multivariate design task.

> The principle helps conduct experiments in graphical design.

Some of those experiments will succeed.

There remain, however, many **other considerations** in the design of statistical graphics — not only of efficiency, but also of **complexity**, structure, density, and even beauty."

— Tufte, Edward, The Visual Display of Quantitative Information



Design is a search problem

Doumont applied to data encoding, which works best? — iterative process of *creating*, *questioning*, *testing*!

Prototypes should emphasize speed over polish.

Get fresh eyes frequently. Invite criticism.

Move from exploring to refining.

© 2021 Scott Spencer / 🌎 https://ssp3nc3r.github.io

scott.spencer@columbia.edu



E HOME **Q** SEARCH



Joe Biden, in Video, Says He Will Be 'More Mindful'



Senate Republicans Go 'Nuclear' to Speed Trump Confirmations

POLITICS ONE-PARTY RULE

Taking the Battle to the States

By HAEYOUN PARK, JEREMY ASHKENAS and MIKE BOSTOCK JAN. 11, 2014

Republicans or Democrats have singleparty control of both the legislature and the governor's office in 36 states, the most in six decades. Lawmakers in these states have been seeking to reshape government policy in recent years, from legalizing same-sex marriage to restricting labor unions. Some of these laws were passed after the rapid rise of single-party control in 2010; others have been in place for years. Below is a look at where states stand on some key issues.





Expanded Medicaid

OH MI





OPENNIS CONFERENCE









scott.spencer@columbia.edu





OPENNIS CONFERENCE



from exploring to explaining

Get our audience(s) to

exploring to explaining, adapting to your audience — remember our goals in communicating with others

pay attention to, understand, (be able to) act upon a maximum of messages, given constraints.





exploring to explaining, adapting to your audience — communicating with mixed audiences

We worked with IR.

We worked with IR. IR stands for Information Resources and is a new department.

We worked with the recently launched Information Resources (IR) department to ...



References and scott.spencer@columbia.edu



exploring to explaining, adapting to your audience — first, optimize encodings for a communication purpose



Position common scale

more effective





less effective

categorical



© 2021 Scott Spencer / 💭 https://ssp3nc3r.github.io

🙊 scott.spencer@columbia.edu





less effective

Once data-to-visual encodings have been optimized for showing the intended comparison or trend of interest to our audience, we should generally *adapt it to our audience by explaining*, not by changing optimized encodings.

exploring to explaining, adapting to your audience — then change annotations, not optimized encodings





exploring to explaining, titles, as an overall graphics annotation, should explain the point of the graphics



Spencer, Scott. Example 3 "Maximizing Game Attendance." In Data *in Wonderland*. <u>https://ssp3nc3r.github.io/data_in_wonderland/</u>

More fans generally pay admission to our games when the chance of winning was near a median of 0.61.

Fans want favorable odds without predicting the outcome.



Sources: Pinnacle betting data, 2016; Retrosheet Gamelogs, 2016

scott.spencer@columbia.edu



exploring to explaining, replacing legends with direct data labeling reduces cognitive load







exploring to explaining, annotations go beyond labels, they can help explain and compare with context

These are tiny paragraphs that explain thing one or thing two.



Rise and Fall of the name **Neil** in the USA Births 1912-2015

Source: data.gov



scott.spencer@columbia.edu



exploring to explaining, annotations as important as data encodings

"The annotation layer is the *most important thing* we do ... otherwise it's a case of 'here it is, you go figure it out'.

— Amanda Cox, the Data Editor at the New York Times



"Annotations are of *vital importance*. Often overlooked, annotations are one of the best ways to make a chart understandable to an audience. Underutilized in many data visualizations, annotations are the ideal way to highlight exactly those things that you, as the creator, want the audience to pay attention to."

— Shirley Wu, *Data Sketches*

"Although our primary focus on creating a visualization is the graphic elements—bars, points, or lines—the **text we include in and around our graphs** is *just as important*."

— Jonathan Schwabish, *Better Data Visualizations*







exploring to explaining, (focus) our audience on explained, visual encodings

remove clutter start with gray

© 2021 Scott Spencer / 💭 https://ssp3nc3r.github.io



References and scott.spencer@columbia.edu

mini-design review *continued* — layering for information hierarchy



00000000			
•	• • • • • • • • • • • •		
		•	
		•	
• • • • •)))))

Orientation	
Shape	
Luminance	
Size	
Hue	
Enclosure	3

design mini-review | layer the graphic, encode visual channels, annotate, and make hierarchies clear



audiences understand graphic complexity — once you explain!



Gaps between Black and white mortality rates for the top 15 causes of death

Note: For non-Hispanic Black and white people in 2018. Rates have been adjusted for age and sex. Source: Centers for Disease Control and Prevention



with guidance

Wezerek, Gus. "Racism's Hidden Toll." The New York Times, August 11, 2020, sec. Opinion. <u>https://www.nytimes.com/</u> interactive/2020/08/11/opinion/us-coronavirus-blackmortality.html.





with guidance

Lu, Denise, and Karen Yourish. "How Did Trump Do in Counties That Backed Him in 2016?" The New York Times, November 11, 2020, sec. Politics. <u>https://www.nytimes.com/</u> interactive/2020/11/09/us/politics/2016-election-trumpcounties.html.

scott.spencer@columbia.edu





U.S. DETAIL 5 The storm can move anywhere inside the cone — and Extreme weather can affect areas far from the cone THE BAHAMAS

with guidance

Cairo, Alberto, and Tala Schlossberg. "Those Hurricane Maps Don't Mean What You Think They Mean." The New York Times, March 10, 2019, sec. Opinion. <u>https://www.nytimes.com/</u> interactive/2019/08/29/opinion/hurricane-dorian-forecastmap.html.







E-commerce jobs are growing fast ...

Employment attributed to electronic shopping firms has doubled in the last five years, outpacing other types of retail.



with guidance

Bebeloff, Robert, and Karl Russell. "How the Growth of E-Commerce Is Shifting Retail Jobs." *The New York Times*, July 6, 2017, sec. Business. <u>https://www.nytimes.com/interactive/</u> 2017/06/business/ecommerce-retail-jobs.html.

... but they are still a small component of overall retail employment ...

Even with the fast growth, the number of online shopping jobs is small compared with department stores, warehouse clubs and grocery stores.



... partly because e-commerce is less labor intensive.

scott.spencer@columbia.edu



Note: Immigration flows from unknown origins not shown. | Source: United Nations Department of Economic and Social Affairs, Population Division

3.4

3.3

14. Mozambique 2.2

15. Philippines 2.0

19. United States 1.7

1.7

20. Ethiopia

24. Morocco

25. Poland

1.6

1.5

9. Italy

10. Pakistan

4. Ukraine

5. Bangladesh 5.5

5.5

with guidance

Porter, Eduardo, and Karl Russell. "Migrants Are on the Rise Around the World, and Myths About Them Are Shaping Attitudes." *The New York Times*, June 20, 2018, sec. Economy. <u>https://www.nytimes.com/interactive/2018/06/20/business/</u> <u>economy/immigration-economic-impact.html</u>.



26.	Iraq	1.5
27.	Spain	1.4
28.	Uzbekistan	1.4
9.	Egypt	1.3
80.	El Salvador	1.2





with guidance

Fairfield, Hannah. "Driving Safety, in Fits and Starts." The New *York Times*, September 12, 2012, sec. Science. <u>https://</u> archive.nytimes.com/www.nytimes.com/interactive/2012/09/17/ science/driving-safety-in-fits-and-starts.html.

empirical study on audience engagement

Haroz, Steve, Robert Kosara, and Steven L. Franconeri. "The Connected Scatterplot for Presenting Paired Time Series." *IEEE* Transactions on Visualization and Computer Graphics 22, no. 9 (September 1, 2016): 2174–86. <u>https://doi.org/10.1109/</u> TVCG.2015.2502587.

(re)design for your audience

redesigns, example — original graphic within government publication explaining part of US economy



All sectors of the U.S. economy contributed to the decrease, led by a decline in private services-producing industries. The decline in first-quarter GDP reflected the response to the spread of COVID-19, as governments issued "stay-at-home" orders in March. This led to rapid changes in production, as businesses and schools switched to remote work or canceled operations, and consumers and businesses canceled, restricted, or redirected their spending. For more information, see "Federal Recovery Programs and BEA Statistics: COVID-19 and Recovery" on the BEA website.



Overall, 17 of 22 industry groups contributed to the first-quarter decline in real GDP. Of the five industry groups that offset the decline in the first-quarter real GDP, agriculture, forestry, fishing, and hunting was the largest contributor, increasing 15.5 percent.

For accommodation and food services, real value added—a measure of an industry's contribution to GDP—decreased 26.8 percent, primarily reflecting a decrease in food services and drinking places.



Finance and insurance decreased 9.0 percent, primarily due to a decrease in insurance carriers and related activities.

Health care and social assistance decreased 7.8 percent, primarily reflecting decreases in ambulatory health care services and in hospitals.

Arts, entertainment and recreation decreased 34.7 percent, primarily reflecting a decrease in performing arts, spectator sports, museums, and related activities.

BEA statistics—including GDP, personal income, the balance of payments, foreign direct investment, the input-output accounts, and economic data for states, local areas, and industries—are available at <u>www.bea.gov</u>. <u>E-mail alerts</u> are also available.

Bureau of Economic Analysis. Gross Domestic Product by Industry: First Quarter 2020. https://www.bea.gov/ sites/default/files/2020-07/gdpind120-fax.pdf.



scott.spencer@columbia.edu



redesigns, example — what's the point of this graphic? Do encodings intuitively show the point? Let's redesign!



Monday, July 6, 2020 Contact: Jeannine Aversa, (301) 278-9003

Gross Domestic Product by Industry: First Quarter 2020

Accommodation and food services; finance and insurance; and health care and social assistance industries were the leading contributors to the 5.0 percent (annual rate) decrease in gross domestic product (GDP) in the first quarter of 2020.

All sectors of the U.S. economy contributed to the decrease, led by a decline in private services-producing industries. The decline in first-quarter GDP reflected the response to the spread of COVID-19, as governments issued "stay-at-home" orders in March. This led to rapid changes in production, as businesses and schools switched to remote -4 work or canceled operations, and consumers and businesses canceled, restricted, or redirected their spending. For more information, see "Federal Recovery Programs and BEA Statistics: COVID-19 and Recovery" on the BEA website.



Overall, 17 of 22 industry groups contributed to the first-quarter decline in real GDP. Of the five industry groups that offset the decline in the first-quarter real GDP, agriculture, forestry, fishing, and hunting was the largest contributor, increasing 15.5 percent.

For accommodation and food services, real value added—a measure of an industry's contribution to GDP—decreased 26.8 percent, primarily reflecting a decrease in food services and drinking places.



Finance and insurance decreased 9.0 percent, primarily due to a decrease in insurance carriers and related activities.

Health care and social assistance decreased 7.8 percent, primarily reflecting decreases in ambulatory health care services and in hospitals.

Arts, entertainment and recreation decreased 34.7 percent, primarily reflecting a decrease in performing arts, spectator sports, museums, and related activities.

BEA statistics—including GDP, personal income, the balance of payments, foreign direct investment, the input-output accounts, and economic data for states, local areas, and industries—are available at www.bea.gov. E-mail alerts are also available.



BEA statistics—including GDP, personal income, the balance of payments, foreign dire and economic data for states, local areas, and industries—are available at www.bea.g







redesigns, example — first possible redesign. Does this redesign more intuitively convey a point?



Real Value Added by Selected Industries



Seasonally adjusted annual rates

BEA statistics—including GDP, personal income, the balance of payments, foreign direct investment, the input-output accounts, and economic data for states, local areas, and industries—are available at <u>www.bea.gov</u>. <u>E-mail alerts</u> are also available.



(Percent change from previous quarter)



Source: U.S. Bureau of Economic Analysis, Seasonally adjusted annual rates



		•					
			-		-		



redesigns, example — second possible redesign. Does this redesign more intuitively convey a point?



Real Value Added by Selected Industries



Seasonally adjusted annual rates

BEA statistics—including GDP, personal income, the balance of payments, foreign direct investment, the input-output accounts, and economic data for states, local areas, and industries—are available at <u>www.bea.gov</u>. <u>E-mail alerts</u> are also available.





redesigns, example — third possible redesign. Does this redesign more intuitively convey a point?



Real Value Added by Selected Industries



Seasonally adjusted annual rates

BEA statistics—including GDP, personal income, the balance of payments, foreign direct investment, the input-output accounts, and economic data for states, local areas, and industries—are available at www.bea.gov. E-mail alerts are also availa2019:Q4



Source: U.S. Bureau of Economic Analysis, Seasonally adjusted annual rates



resources

References

Spencer, Scott. "Visual." Sec. 2 In Data in Wonderland. 2021. https://ssp3nc3r.github.io/ data in wonderland.

Bertini, Enrico, and Moritz Stefaner. "Visual Storytelling w/ Alberto Cairo and Robert Kosara." In Data Stories: A podcast on data visualization with Enrico Bertini and Moritz *Stefaner*. Podcast, April 16, 2014. <u>https://datastori.es/data-stories-35-visual-storytelling-</u> w-alberto-cairo-and-robert-kosara/

Bostock, Michael. "Design Is a Search Problem." Presented at the OpenViz Conference, Boston, MA, 2014. http://www.openvisconf.com/2014/.

Bremer, Nadieh. *Data Sketches A Journey of Imagination, Exploration, and Beautiful Data* Visualizations. Milton, UNITED KINGDOM: A K Peters/CRC Press, 2021.

- **Corum**, Jonathan. "See, Think, Design, Produce 3." 13pt Information Design, March 28, 2016. http://style.org/stdp3/.
- ———. "Design for an audience" 13pt Information Design, April 26, 2018. <u>http://</u> style.org/ku/.
- **Doumont**, Jean-Luc. "Fundamentals." In *Trees, Maps, and Theorems. Effective* Communication for Rational Minds. Principiæ, 2009.
- **Schwabish**, Jonathan A. *Better Data Visualizations: A Guide for Scholars, Researchers, and* Wonks. New York: Columbia University Press, 2021.
- Tufte, Edward R. The Visual Display of Quantitative Information. Second. Graphics Press, 2001.
 - **Ware**, Colin. *Information Visualization: Perception for Design*. Fourth. Philadelphia: Elsevier, Inc, 2020.





extra

explain for audience, be open to new approaches for visual communication when designing for an audience



— Meeks, Elijah. "If Data Visualization Is So Hot, Why Are People Leaving?" Blog. Medium, March 21, 2017. https://medium.com/@Elijah_Meeks/why-people-leave-their-data-viz-jobs-be1a7ab5dddc.





redesigns, advice from Jonathan Corum, Science Editor at the New York Times



By Jonathan Corum April 26, 2018

Last week I gave a workshop and talk at SUND, the University of Copenhagen's Faculty of Health Sciences.

Here's a lightly edited transcript of the talk:

Thank you very much for being here. Thank you for the

Find the visual idea Translate Tell a visual story Focus attention, don't scatter it Show the content, not the frame Show the content, not the table Be consistent What can you remove? Reference the real world Connect images and data Explain why Provide context Build a sequence Show movement and change Encourage visual comparisons More labels, fewer legends Annotate