



University of Sheffield

**DATA VISUALISATION COMMUNITY GROUP OF PRACTICE**

---

# **Translating Complex Research into Effective Visuals**

*From Research Design to Research Output -  
The Social Science Perspective*

**Dr. Jacob L. Macdonald**

[j.macdonald@sheffield.ac.uk](mailto:j.macdonald@sheffield.ac.uk)

Oct. 11, 2022

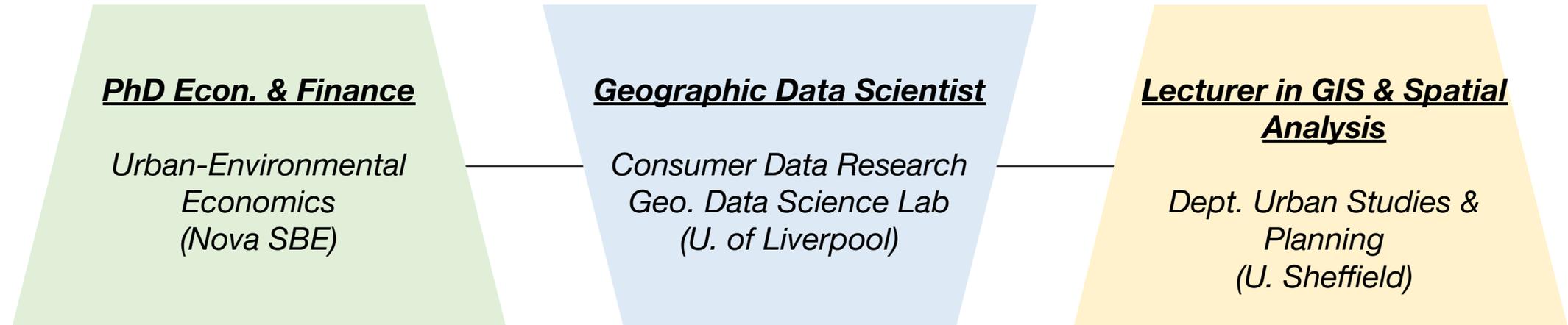


The  
University  
Of  
Sheffield.

# Background & Research Context

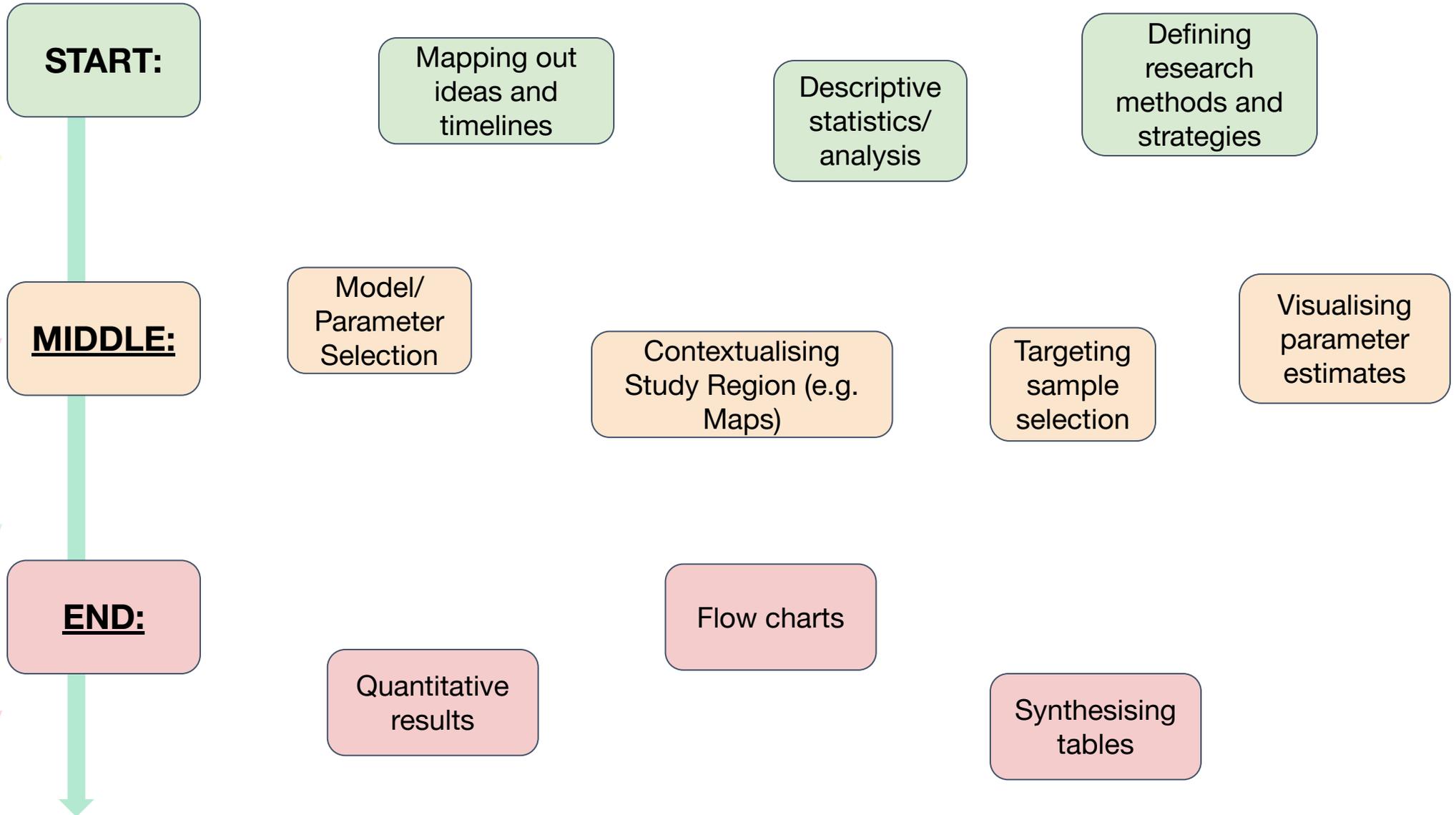
Social sciences perspective –

Research questions and data on human behaviour, economic and social processes, the natural and the built environment.



{Big/ Spatial/ Secondary/ Open} data for research and analysis – mainly focused on valuing and quantifying urban built and natural environmental features and policies.

# Visualisations Across the Research Lifecycle



# Importance of Good Visualisations

---

**Good data visualisations** can be used as a call to action - highlighting the extent of a problem or issue, helping supporting *everybody* in understanding the research.

**Bad data visualisations** can be confusing and get in the way of the discussion, distracting from the issue.

# Importance of Good Visualisations

*E.g.*

Take how information was conveyed during the recent **COVID-19** pandemic - the public really found the visuals of the data to be important.

nature index

Home News Current Index Annual tables Supplements Client services About

Home > News > Data visualizations are key to COVID-19 communication, but we still don't understand their impact

NEWS · 18 AUGUST 2020

## Data visualizations are key to COVID-19 communication, but we still don't understand their impact

The language of data visualization has become commonplace, but their influence on public opinion and behaviour is unclear.

Helen Kennedy

XD IDEAS / PERSPECTIVES / LEADERSHIP INSIGHTS

## Communicating COVID-19 Complexity Through Data Visualization

Linn Vizard  
Jun 5, 2020

UCL Home > UCL News > Analysis: data visualisation expert on what's wrong with the UK Government's coronavirus charts

## Analysis: data visualisation expert on what's wrong with the UK Government's coronavirus charts

3 November 2020

Professor James Cheshire (UCL Geography) explains what's wrong with the way the data is presented at the UK Government's Covid-19 briefings, as well as how to improve it.

Follow us

f i y t n

Tweets from @uclnews

# ... and the Importance of Good Tables

And not just the visualisations that need to be considered - we should also take into account having well presented data tables and results.

Presenting numbers and quantitative data in a paper can be equally as important. When doing so, for example, we should have:

- Well labeled variables, font choices, spacing
- Clean and appropriate scientific notation (digits)
- Appropriate emphasis on certain rows or columns
- Reduced clutter

*An overall **focusing on telling a story** with the output to an accessible audience.*

**Graphics and statistics for cardiology: designing effective tables for presentation and publication**

(Boers, 2018)

<https://doi.org/10.1136/heartjnl-2017-311581>

# ... and the Importance of Good Tables

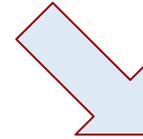
**Table ##** Changes (mean change or change in proportion followed by 95% CI) in lifestyle, medical risk factors, cardioprotective medication and patient-reported outcome measures in cardiovascular patients between the IA, EOP and 1-year assessments

	IA (n=549)	EOP (n=549)	Change	IA (n=231)	1 year (n=231)	Change
Current smoking (%)	9.3%	7.8%	-1.5% (-3.5%, 0.5%); p=0.16	8.8%	8.3%	-0.4% (-3.7%, 2.9%); p=1.00
Fruit and vegetable: ≥5 portions/day (%)	35.6%	47.9%	12.2% (7.7%, 16.8%); p<0.001	34.2%	45.8%	11.6% (4.2%, 18.9%); p=0.002
Fish: >20 g/day (%)	75.2%	84.4%	9.1% (5.2%, 13.1%); p<0.001	75.7%	83.6%	8.0% (1.7%, 14.2%); p=0.01
Mediterranean Diet Score mean (SD)	7.5 (2.2)	8.5 (1.9)	1.0 (0.9, 1.1); p<0.001	7.2 (2.3)	8.3 (2.0)	1.2 (0.9, 1.4); p<0.001
Physical activity: ≥5 times/week, ≥30 min (%)	17.0%	56.8%	39.8% (34.5%, 45.0%); p<0.001	16.1%	47.9%	31.7% (24.2%, 39.3%); p<0.001
Estimated METs maximum mean (SD)	7.6 (2.0)	8.5 (2.2)	0.9 (0.8, 1.1); p<0.001	7.6 (1.8)	8.5 (2.0)	0.9 (0.7, 1.2); p<0.001
Weight† kg mean (SD)	84.0 (14.0)	83.8 (13.7)	-0.2 (-0.6, 0.1); p=0.43	84.4 (14.6)	84.1 (14.3)	-0.3 (-1.0, 0.3); p=0.35
Waist circumference, cm mean (SD)	97.0 (12.4)	96.4 (11.9)	-0.6 (-0.9, -0.3); p<0.001	98.8 (10.9)	97.9 (10.8)	-0.8 (-1.6, -0.1); p=0.03
BP <140/90 mm Hg (%)	70.3%	85.6%	15.4% (11.0%, 19.7%); p<0.001	69.9%	79.0%	9.2% (2.2%, 16.2%); p=0.01
Total cholesterol <4 mmol/L (%)	58.1%	64.5%	6.4% (1.9%, 10.9%); p=0.005	58.8%	57.3%	-1.5% (-8.8%, 5.7%); p=0.77
LDL-cholesterol <2 mmol/L (%)	48.2%	54.5%	6.3% (2.0%, 10.6%); p<0.004	49.5%	49.0%	0.5% (-8.3%, 7.3%); p=1.00
HbA1c mmol/mol mean (SD)*	56.6 (15.0)	57.0 (14.1)	0.4 (-2.8, 3.6); p=0.81	53.7 (15.4)	56.6 (13.5)	2.9 (-3.2, 9.0); p=0.16
Antiplatelet therapy (%)	91.8%	91.6%	-0.2% (-1.7%, 1.3%); p=1.00	92.1%	89.1%	-3.1% (-6.6%, 0.4%); p=0.09
Statins (%)	90.2%	92.2%	2.0% (-0.2%, 4.2%); p=0.08	92.2%	94.4%	2.2% (-1.5%, 5.9%); p=0.30
ACE inhibitors/ARBs (%)	69.7%	76.2%	6.5% (3.3%, 9.6%); p<0.001	71.2%	73.8%	2.6% (-2.0%, 7.2%); p=0.31
Beta blockers (%)	71.3%	68.7%	-2.6% (-5.2%, 0.0%); p=0.05	72.9%	70.3%	-2.6% (-7.1%, 1.8%); p=0.29
Calcium channel blockers (%)	18.2%	21.9%	3.7% (1.2%, 6.2%); p=0.002	17.9%	17.5%	-0.4% (-4.4%, 3.5%); p=1.00
Diuretics (%)	22.0%	20.7%	-1.3% (-3.9%, 1.3%); p=0.36	20.5%	19.7%	-0.9% (-5.3%, 3.6%); p=0.83
HADS-Anxiety median (IQR)	5 (2, 8)	5 (2, 8)	0 (0, 0); p=0.55	5 (2, 9)	5 (2, 7)	-1 (-1, 0); p=0.06
% HADS-Anxiety >8	27.8	26.3	-1.5% (-8.4%, 5.3%); p=0.76	30.4	19.6	-10.7% (-19.9%, -1.5%); p=0.02
HADS-Depression median (IQR)	3 (2, 7)	2 (1, 6)	-1 (-1, 0); p<0.001	4 (2, 7)	3 (1, 6)	-1 (-1, 0); p=0.002
% HADS-Depression >8	23.2	11.6	-11.6% (-17.7%, -5.5%); p<0.001	24.1	16.1	-8.0% (-17.2%, 1.1%); p=0.09
EQ-VAS (IQR)	65 (50, 77)	73 (60, 85)	5 (0, 10); p<0.001	60 (50, 80)	75 (60, 80)	10 (0, 10); p<0.001
Dartmouth COOP median (IQR)	22 (17, 27)	19 (15, 24)	-2 (-2, -1); p<0.001	23 (17, 27)	21 (17, 25)	-1 (-3, 0); p=0.008

\*In those with known diabetes or newly diagnosed diabetes at the IA.

†In those with BMI >25 kg/m<sup>2</sup> at IA

ARBs, angiotensin receptor blockers; BMI, body mass index; EOP, end of programme; HADS, Hospital Anxiety and Depression Scale; IA, initial assessment; EQ-VAS, Euroqol Visual Analogue Scale.



	baseline		program end		change (n=549)			
					mean (95%CI)		p-value	
Current smoking (%)	9		8		-3,5	-1,5	0,5	0,16
Fruit and vegetable: ≥ 5 portions/day (%)	36		48		7,7	12,2	16,8	<0,001
Mediterranean Diet Score mean (SD)	7,5	2,2	8,5	1,9	0,9	1	1,1	<0,001
Estimated METs maximum mean (SD)	7,6	2,0	8,5	2,2	0,8	0,9	1,1	<0,001
HADS -Anxiety median (IQR)	2	5	8	2	5	8	0	0,55
% with score >8	28		26		-8,4	-1,5	5,3	0,76
-Depression median (IQR)	2	3	7	2	2	6	-1	<0,001
EQ-VAS (IQR)	50	65	77	60	73	85	0	5
Dartmouth COOP median (IQR)	17	22	27	15	19	24	-2	-2
							-1	<0,001

Graphics and statistics for cardiology: designing effective tables for presentation and publication

(Boers, 2018)

<https://doi.org/10.1136/heartjnl-2017-311581>

# Consider the Elements of Design

They all apply in scientific graphics and figures as well:

Line

Colour

Shape

Space

Texture

Typography

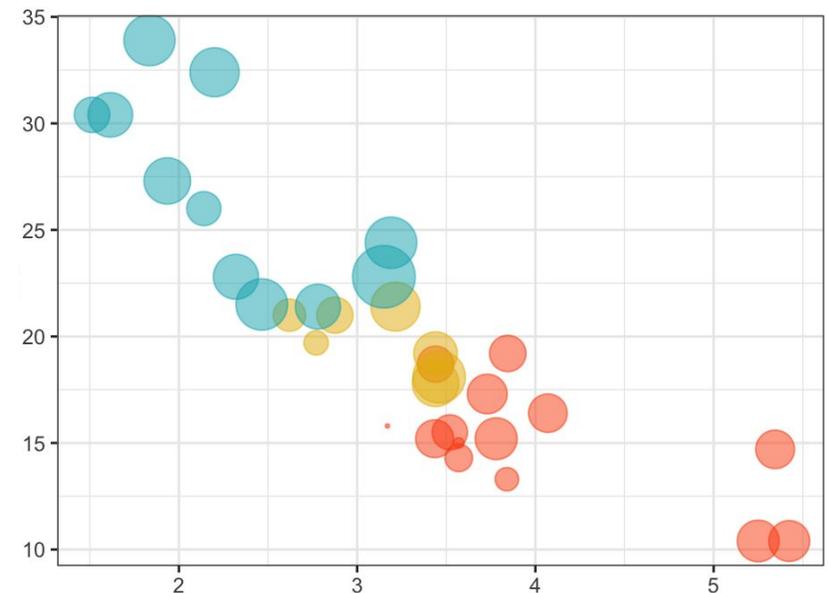
Scale (Size)

Emphasis/ Contrast

Dominance/ Hierarchy

We can use all of these elements in different ways to transmit different ideas, quantitative values, research outputs and any other graphic visuals we wish to create.

**Combinations of these different elements can help us to present multidimensional effects in 2D.**



# How to Convey Information

<b>Guideline 1:</b> Create the simplest graph that conveys the information you want to convey.	<b>Guideline 6:</b> Plot overlapping points in a way that density differences become apparent in scatter plots.
<b>Guideline 2:</b> Consider the type of encoding object and attribute used to create a plot.	<b>Guideline 7:</b> Use lines when connecting sequential data in time-series plots.
<b>Guideline 3:</b> Focus on visualizing patterns or on visualizing details, depending on the purpose of the plot.	<b>Guideline 8:</b> Aggregate larger datasets in meaningful ways.
<b>Guideline 4:</b> Select meaningful axis ranges.	<b>Guideline 9:</b> Keep axis ranges as similar as possible to compare variables.
<b>Guideline 5:</b> Data transformations and carefully chosen graph aspect ratios can be used to emphasize rates of change for time-series data.	<b>Guideline 10:</b> Select an appropriate color scheme based on the type of data.

**Ten guidelines for for effective data visualization in scientific publications**

(Kelleher & Wagener, 2011)

<https://doi.org/10.1016/j.envsoft.2010.12.006>

# A Good Scientific Visual



**Tells a story**

**Is accessible, well presented**

**Maintains the 'human' perspective**

**Keeps it simple**

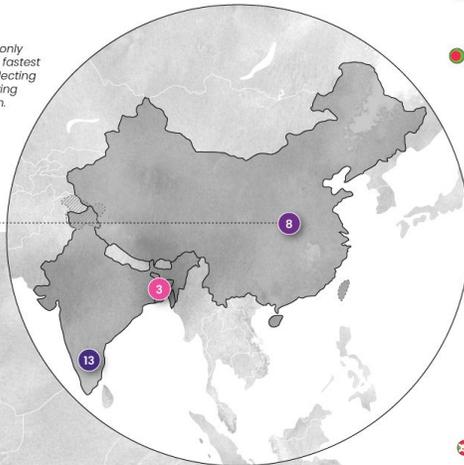
Top 20 Ranked

# The Fastest Growing Cities Worldwide



By 2025 the world's population will sit at over **8.1 billion people**. While some regions, like Europe, will experience slow growth, many up-and-coming cities are set to see rapid population expansions by 2025.

China is home to only one of the world's fastest growing cities, reflecting the country's slowing population growth.

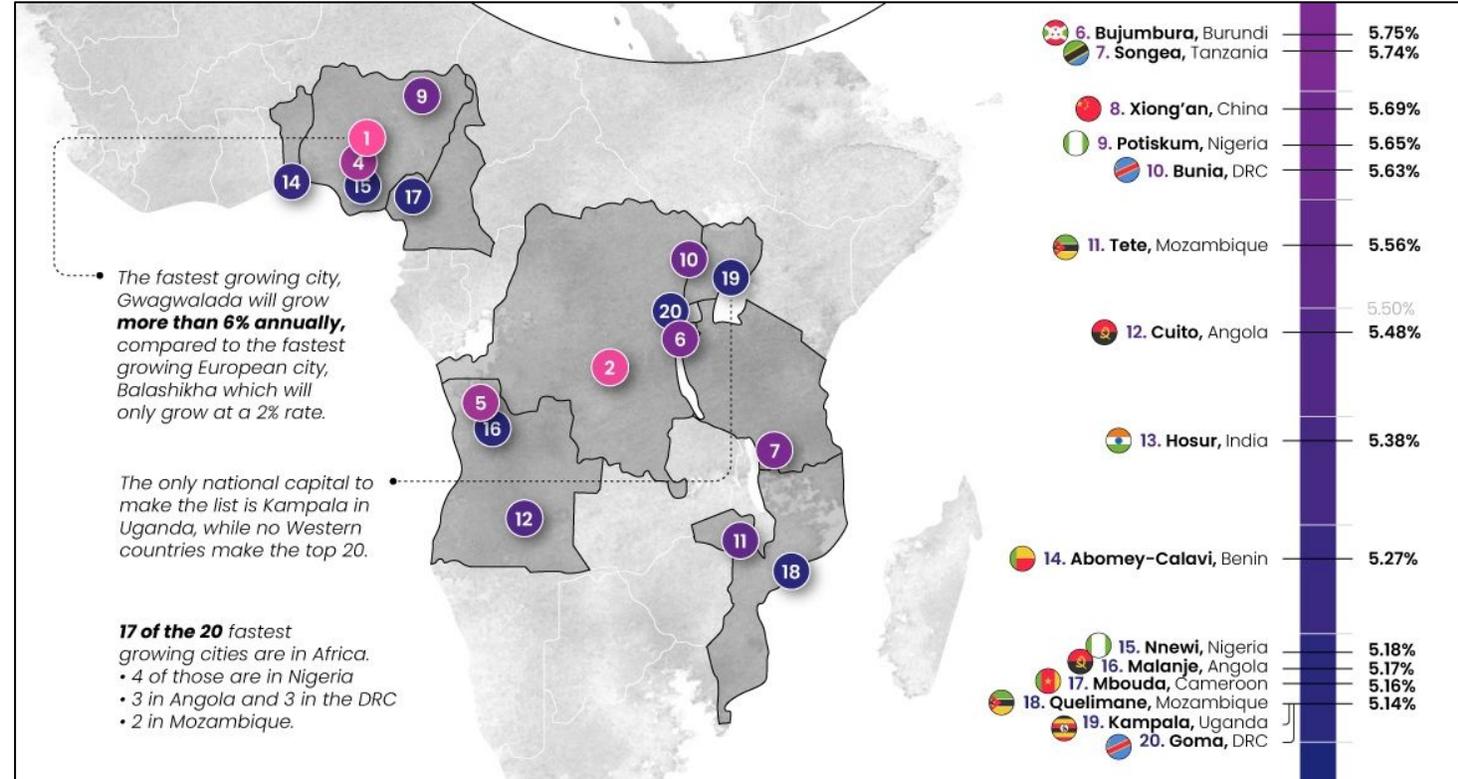
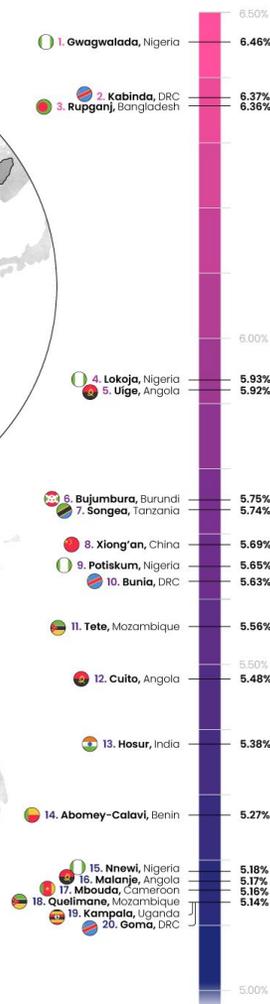


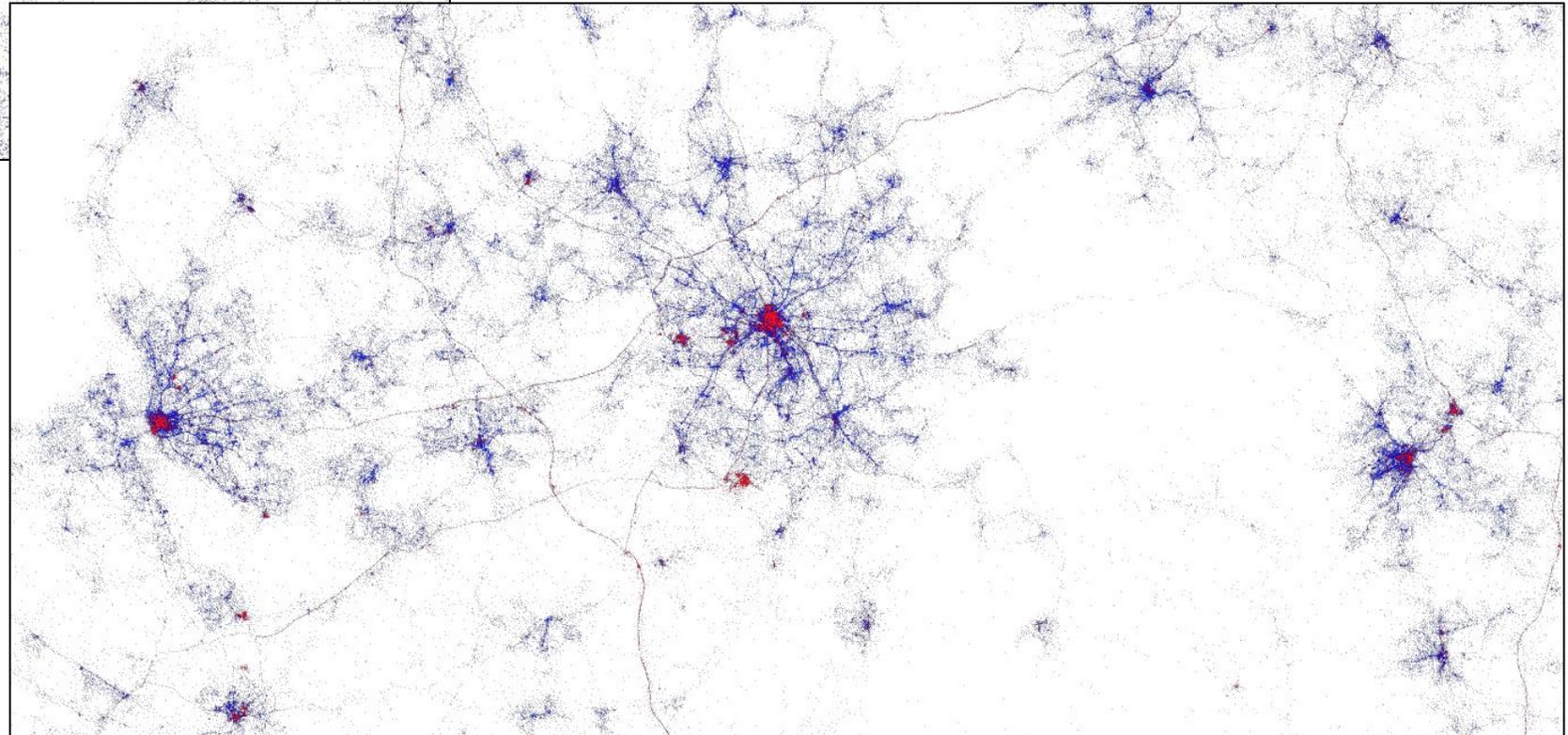
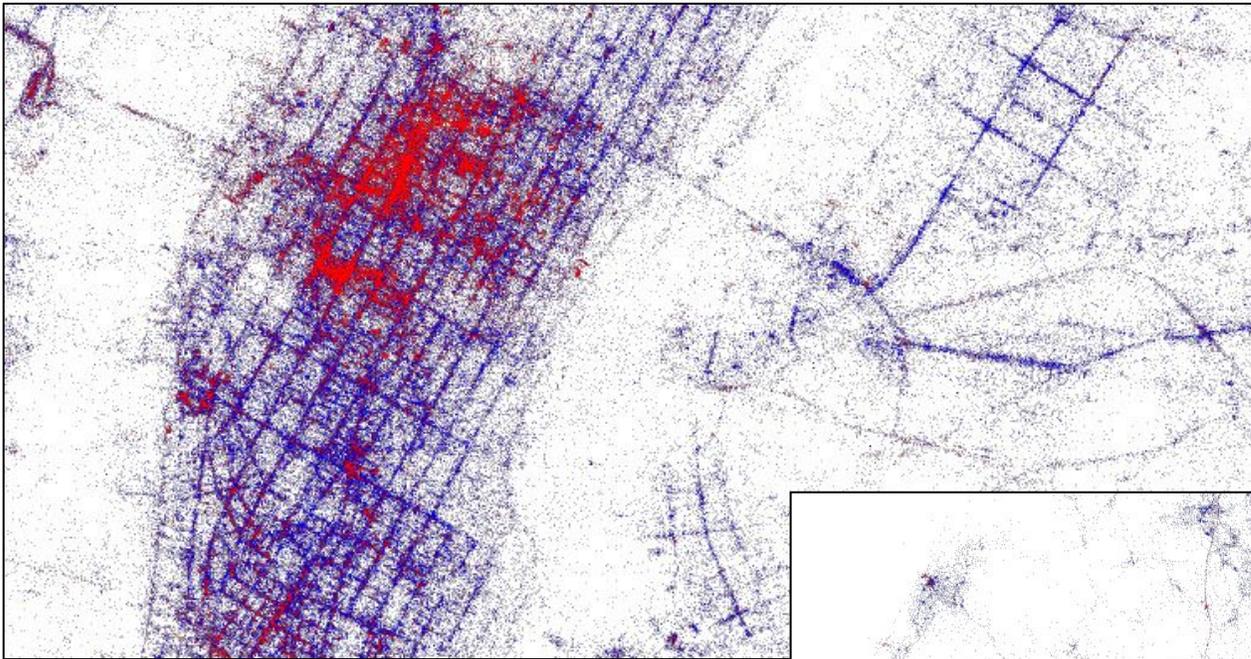
The fastest growing city, Gwagwalada will grow **more than 6% annually**, compared to the fastest growing European city, Balashikha which will only grow at a 2% rate.

The only national capital to make the list is Kampala in Uganda, while no Western countries make the top 20.

**17 of the 20** fastest growing cities are in Africa.  
 • 4 of those are in Nigeria  
 • 3 in Angola and 3 in the DRC  
 • 2 in Mozambique.

Annual growth rate 2020p-2025p



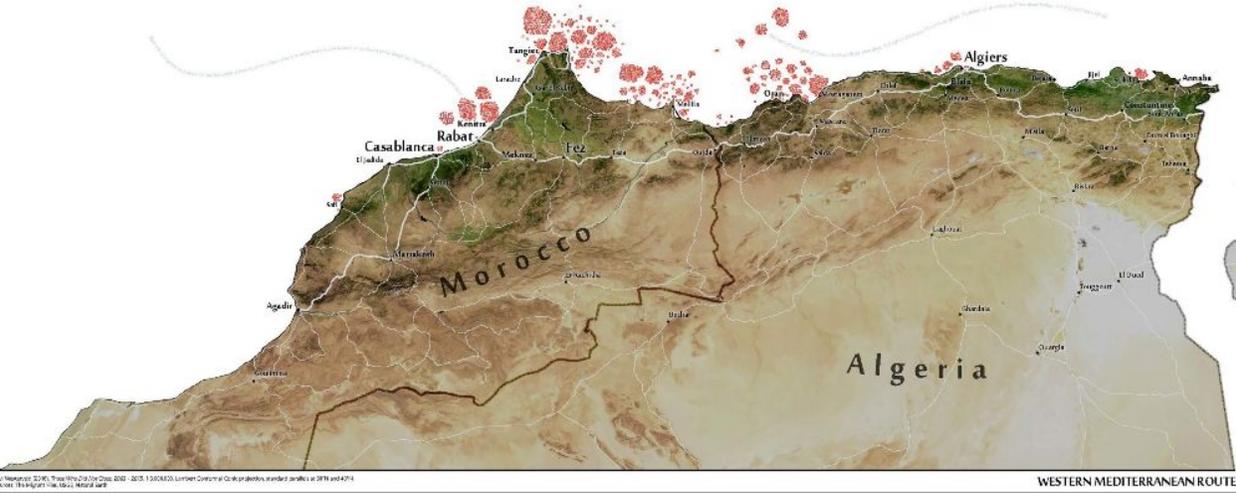
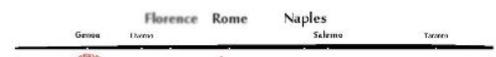
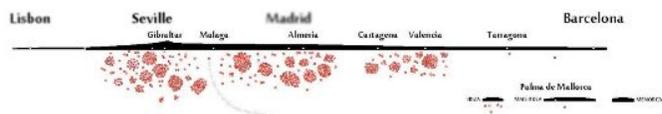


# THOSE WHO DID NOT CROSS 2005 -2015

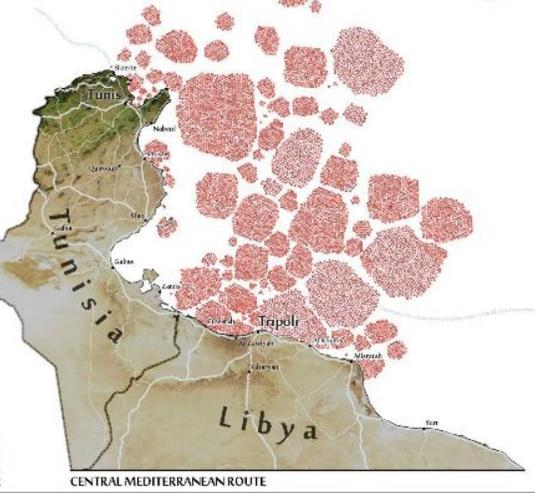
(December 3, 2011) A 29-year-old pregnant woman died from drinking seawater after the boat she embarked on in Libya went adrift in the Mediterranean Sea for 16 days. From 2005 to 2015, over 16,000 other individuals were reported dead or missing as they tried to reach European shores fleeing conflict and instability in Africa and the Middle East.

Libya has become a popular starting point for many journeys, with people traffickers exploiting the country's power vacuum and increasing lawlessness. The relatively short distance to Lampedusa encourages more people to risk the journey.

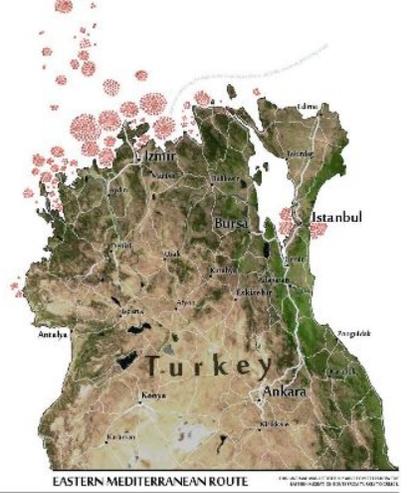
Each red dot on this map shows where a person went missing or died on the Western, Central, and Eastern Mediterranean routes to Europe.



WESTERN MEDITERRANEAN ROUTE



CENTRAL MEDITERRANEAN ROUTE



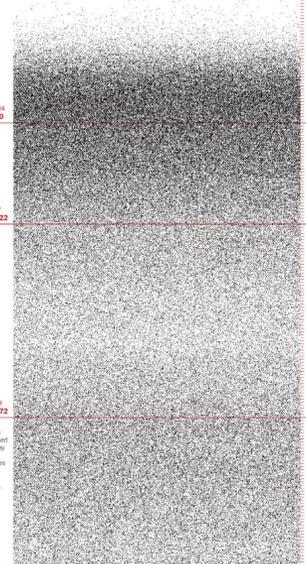
EASTERN MEDITERRANEAN ROUTE

U.S. VIRUS DEATHS NEARING 500,000 IN JUST ONE YEAR MORE THAN IN 3 YEARS

By Julie Bosman
CHICAGO — A nation numbed by misery and loss is confronting a number that still has the power to shock: 500,000.

The Toll: America Approaches Half a Million Covid Deaths

Feb. 20, 2021: first report of a U.S. death in Washington State



Garland Faces Resurgent Peril Of Extremism

Oklahoma City Attack Shaped His Views

By Mark Leibovich
WASHINGTON — Judge Merrick B. Garland shows no signs of being intimidated by the resurgent threat of extremism.

STORMS EXPOSING A NATION PRIMED FOR CATASTROPHE FOR CATASTROPHE

CLIMATE CHANGE WRAITH Unprepared for Threats Facing Power Grids, Water and Roads

By Christopher Flavelle, Brad Plumer and Hiroko Tabuchi
Even as Texas struggled to restore electricity and water over the past week, signs of the risk posed by increasingly extreme weather to America's aging infrastructure were cropping up across the country.

JOB LOSSES SOAR; U.S. VIRUS CASES TOP WORLD



Since the coronavirus descended on Brooklyn Hospital Center three weeks ago, the staff has handled over 800 potential cases.

New Data Shows Staggering Toll of Outbreak

By Ben Gompertz
The coronavirus outbreak in the United States is now the largest in the world, with more than 1 million cases and nearly 200,000 deaths.

Unforced Posts Force Scramble By Washington

By Jennifer Steinhilber and Alan Kohan
WASHINGTON — Of the 75 million posts on Facebook that were removed last week, the vast majority were unforced posts.

Online Class With No Way To Get There

By Nadine Stuenkel
Allison Phillips was excited about trying an online class for the first time, but she found it frustrating.

Courage at a Brooklyn Hospital, At the Front of an Invisible War

By Robert Fink
Brooklyn Hospital Center is a frontline in the fight against the coronavirus, with staff working long hours in a high-stress environment.

No Crowd, but I'll Take You Out to the Ballgame

By Dan Barry
Professional baseball games are being played in empty stadiums, but fans are still showing up to support their teams.

A Fan Writes a Fantasy For Opening Day

From the material for the first baseball game, a fan has written a fantasy story about the opening of the season.

Labeling Begins by Workers

Workers are beginning to label their own contributions to the fight against the coronavirus, marking a shift in the narrative.

Russia Fears But Can't Quit Open Internet

By Anton Troianovski
MOSCOW — Margarita Simonyeva, the editor in chief of the Kremlin-controlled RT television network, recently called for a return to a free internet.

Lack of Birth Control Deepens Women's Burden in Venezuela

By Julie Turkewitz and Isayen Herrera
SAN DIEGO DE LOS ANGELES, Venezuela — The moment Johanna Guzman, 25, discovered she was going to have her sixth child, she knew she was in trouble.

Israel Funds Vaccines for Syria

TRACKING AN OUTBREAK 4-9
ISRAELI SECURITY SERVICES have agreed to finance a supply of Russian-made Covid-19 vaccines for Syria, an official said.

Osaka Wins Australian Open

SPORTS 35-37
NAOMI OSAKA, 23, won a 4-for-4 Grand Slam final after a rain-shortened victory over Jennifer Brady, 25, who was in her first Slam final.

Boredom Is Making Us Buy

SUNDAY BUSINESS
Yet another pandemic side effect: consumers who are eager to make online purchases, and not just of groceries.

Brushing Aside a Slaughter

INTERNATIONAL AIR
KATIE ENGLERT
A slaughterhouse in Iowa is being shut down after a worker was found to have contracted COVID-19.



Armed With Sewing Machines

Armed with sewing machines, a group of women in a small town in Iowa are making face masks to help combat the coronavirus.

Malaria Is Indicted in U.S.

NATURAL AZLES
A malaria parasite has been indicted in the United States for its role in causing the disease.

New Focus on Health System

WEEKEND BATS 11-10
The coronavirus pandemic has led to a renewed focus on the health system's ability to handle future outbreaks.

For Fine-Line Paleontologists

For fine-line paleontologists, the coronavirus pandemic has led to a renewed focus on the importance of their work.

It's Virtually Perfect

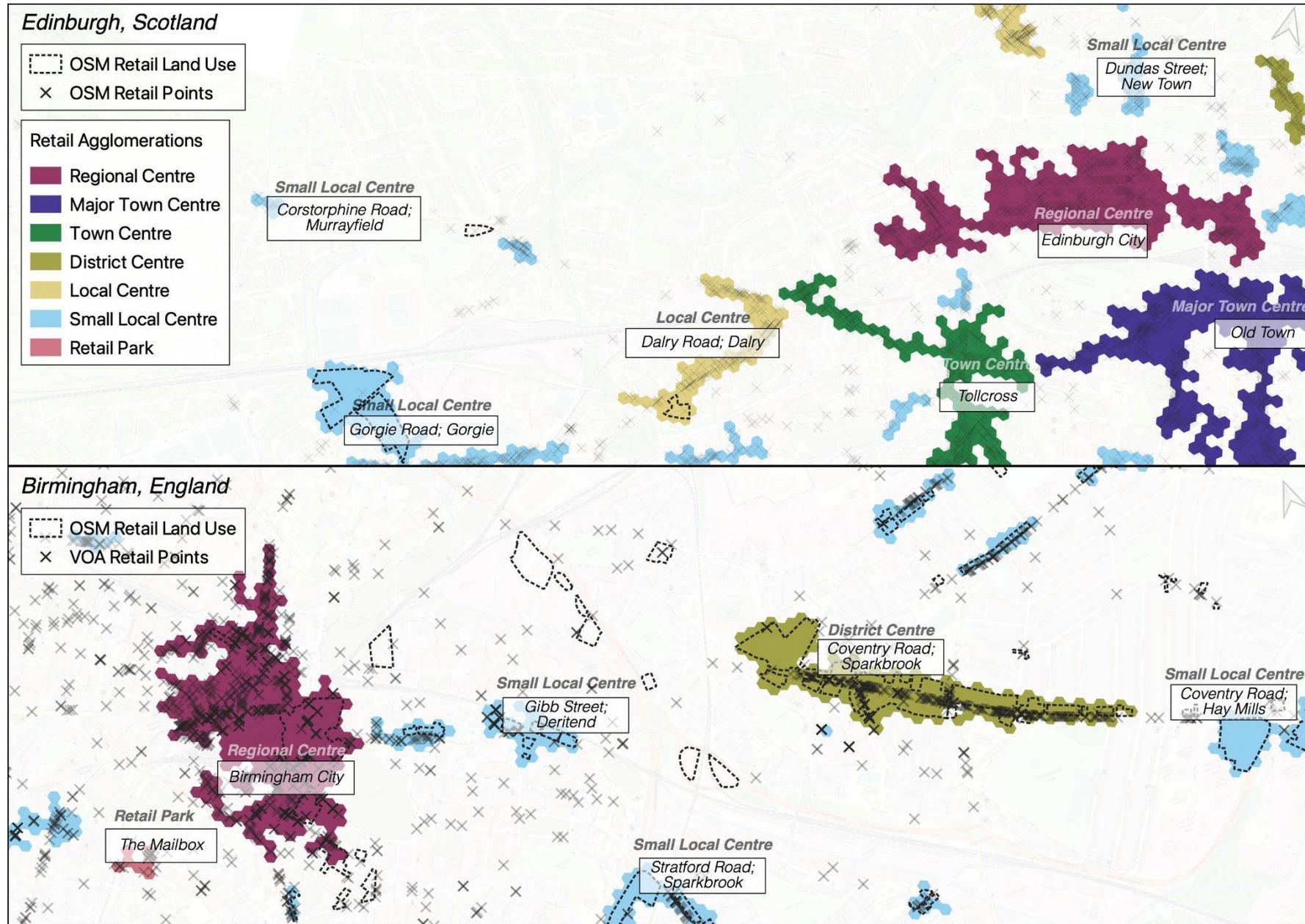
It's virtually perfect: a new online platform for virtual reality experiences has been launched.

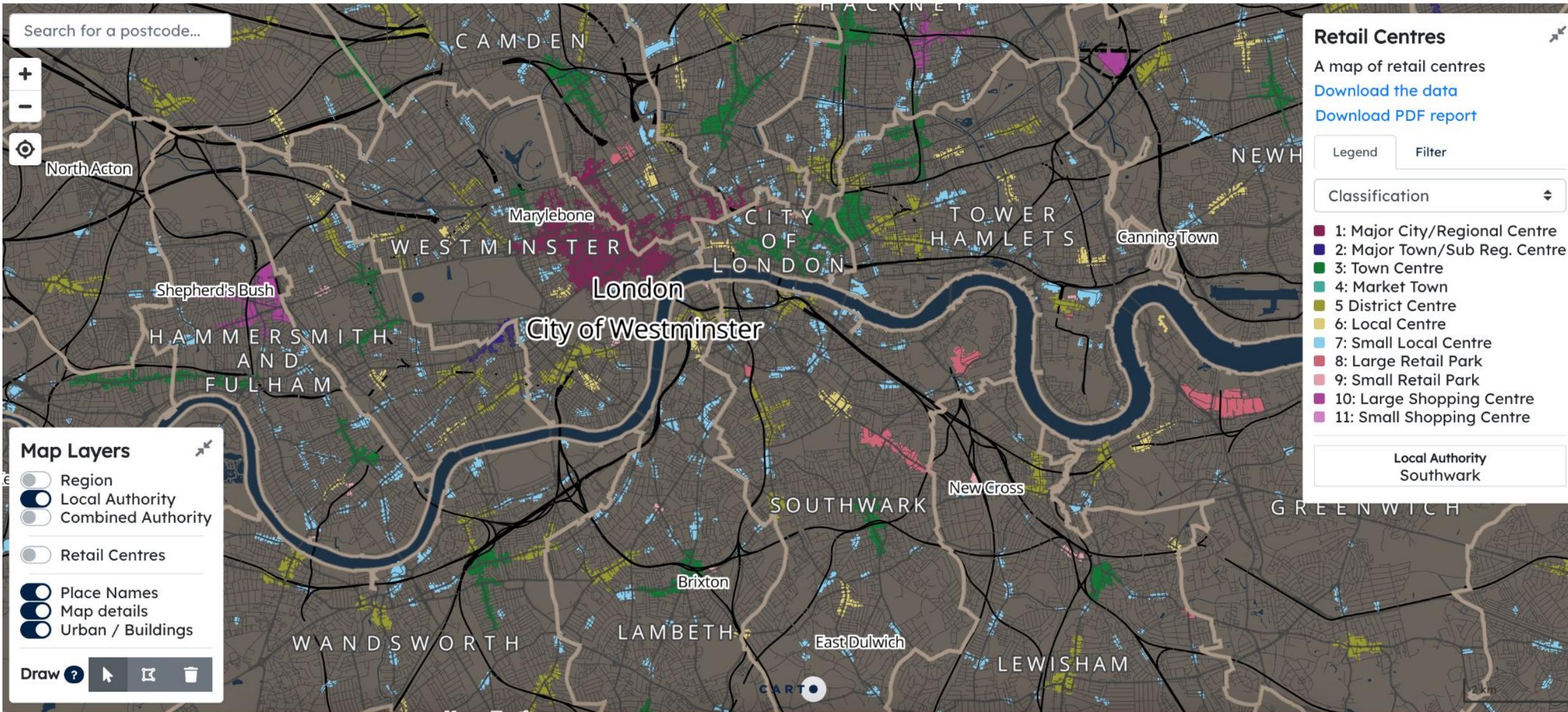
Chancellor of U.S. History

Chancellor of U.S. History: A new role has been created to oversee the nation's historical sites and museums.

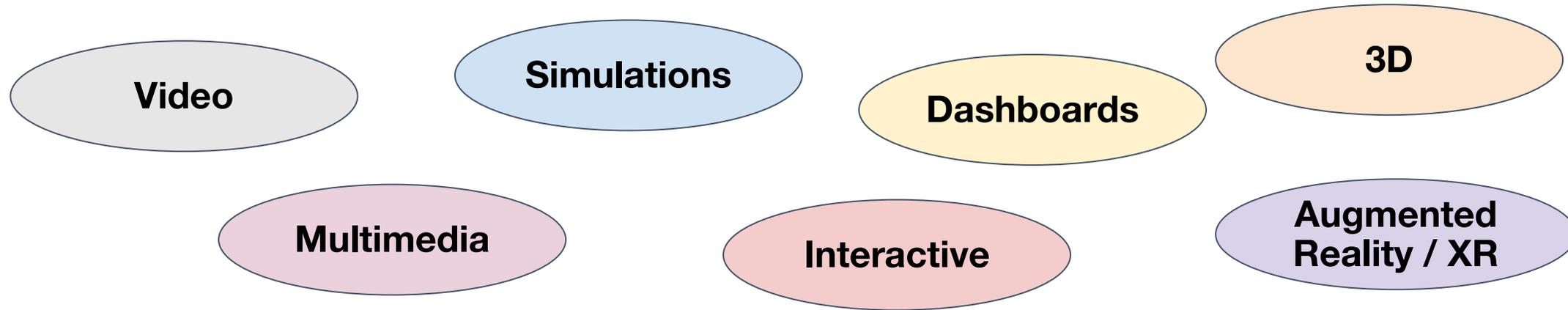
Gabriel's Dean of Dishes

Gabriel's Dean of Dishes: A new position has been created to oversee the nation's culinary arts education.





# Output Types and Formats



**Still**, a well designed 2D (printable) representation is always good to have - for power points; presentations; academic articles; etc.

Not every computer or display can showcase every type of visual.

Having a target set of output formats considered beforehand can go a long way towards helping you present your research and discussion.

Having key figures and graphs already developed which are well formatted, properly labeled and with an accessible format, means that you can quickly put presentations together and share ongoing project highlights.

# Accessible Visuals

Making sure that your graphics are accessible is important - especially now that much of our resources will be presented online.

- Accessible and distinguishable **colour** palettes.
- Appropriate **contrast** in boundaries, shapes, colours.
- Appropriate and **legible font** size (preferable sans serif/ without flourishes).
- Provide some **Alt-text** for your visuals so that you can best describe the visual outputs based on your research knowledge.

You know best how to describe your research and outputs.

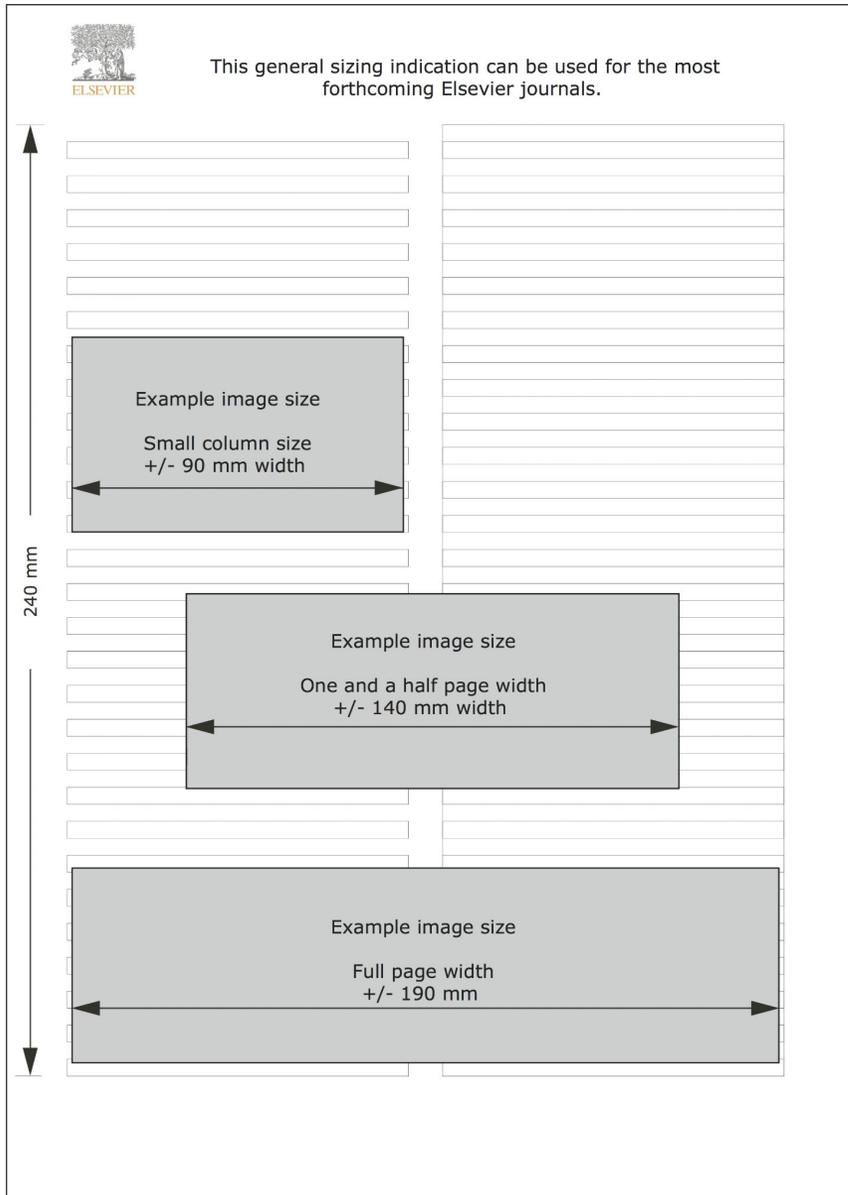
**Accessible Color Sequences for Data Visualization**

(Petroff, 2021)

<https://arxiv.org/abs/2107.02270>

Consider different perspectives and how a range of people with varying visual abilities may be accessing your works and outputs.

# Journal Guides - Visuals for publishing



If you know you'll be submitting this work to an academic journal, it doesn't hurt to preemptively take into account their technical limitations (for print media).

Can save a lot of time and energy towards the end of the project.

**E.g.** Checking the [Elsevier](#) or [Sage guidelines](#) around what is required.

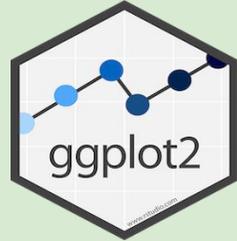
- 300 DPI
- Printed in BW
- Sans serif fonts
- Format (Tiff; Jpeg; Eps)

# The (Open Source) Tools and Resources

## Softwares/ Libraries

Python

Matplotlib  
Seaborn  
Bokeh  
Plotly



Data Studio

Poly**maps**

**Chartist.js**

**RAW**Graphs

Leaflet The logo for Leaflet, featuring the word 'Leaflet' in a cursive font and a green leaf icon.

The logo for Blender, featuring an orange and blue stylized 'b' icon and the word 'blender' in a bold, lowercase font.

## Open Data

Local  
NGOs

Municipalities

Stats.  
Agencies

Open Data  
Platforms

National  
Geographic  
Surveys

Open  
Street  
Map

Government;  
Ministries

## Community & Resources

Alan Turing Network

<https://www.turing.ac.uk/research/interest-groups/visualization>

Sheffield AT  
Network  
Interest Group

Data Vis. Journals

<https://lib.guides.umd.edu/datavisualization/publish>

(US) Urban Inst.  
Style Guide:

<https://urbaninstitute.github.io/graphics-styleguide/>

Sheffield Data  
Visualisation  
Community

Visualising Data Blog:

<https://www.visualisingdata.com/resources/>



*j.macdonald@sheffield.ac.uk*



@jacobmac02

N