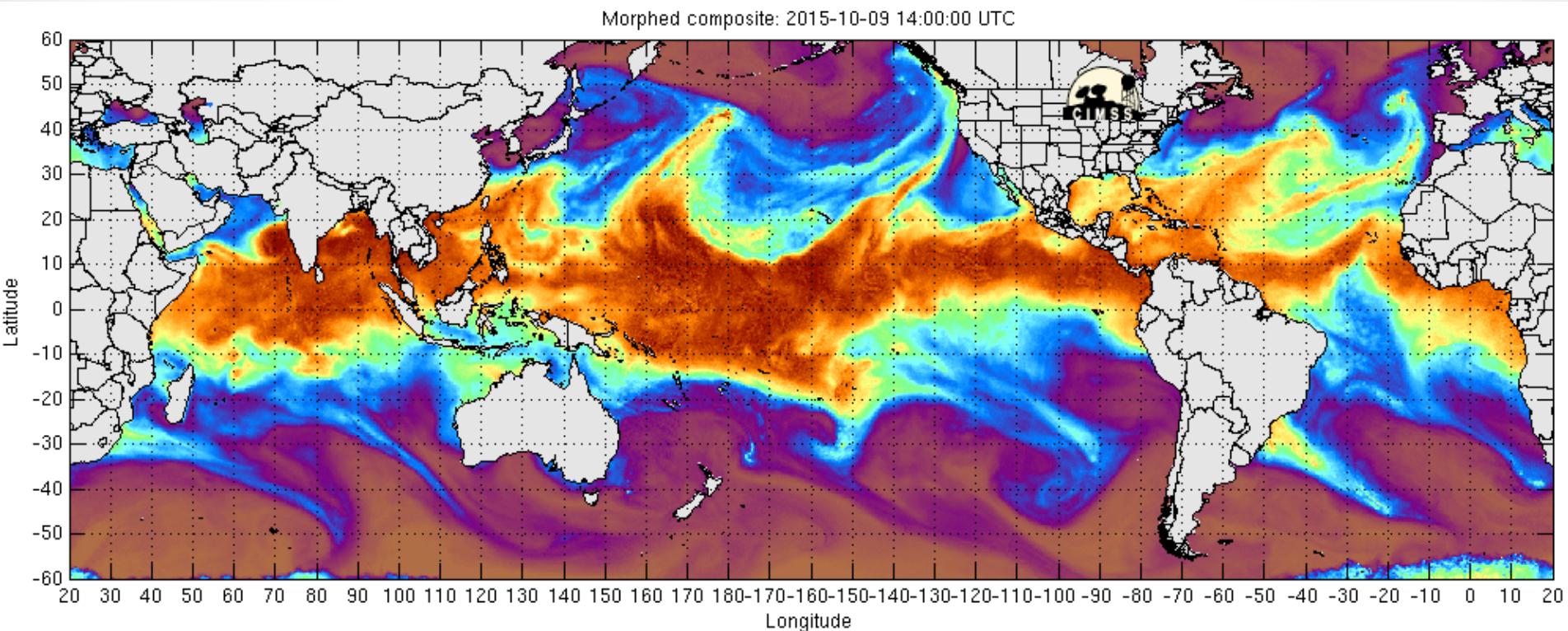
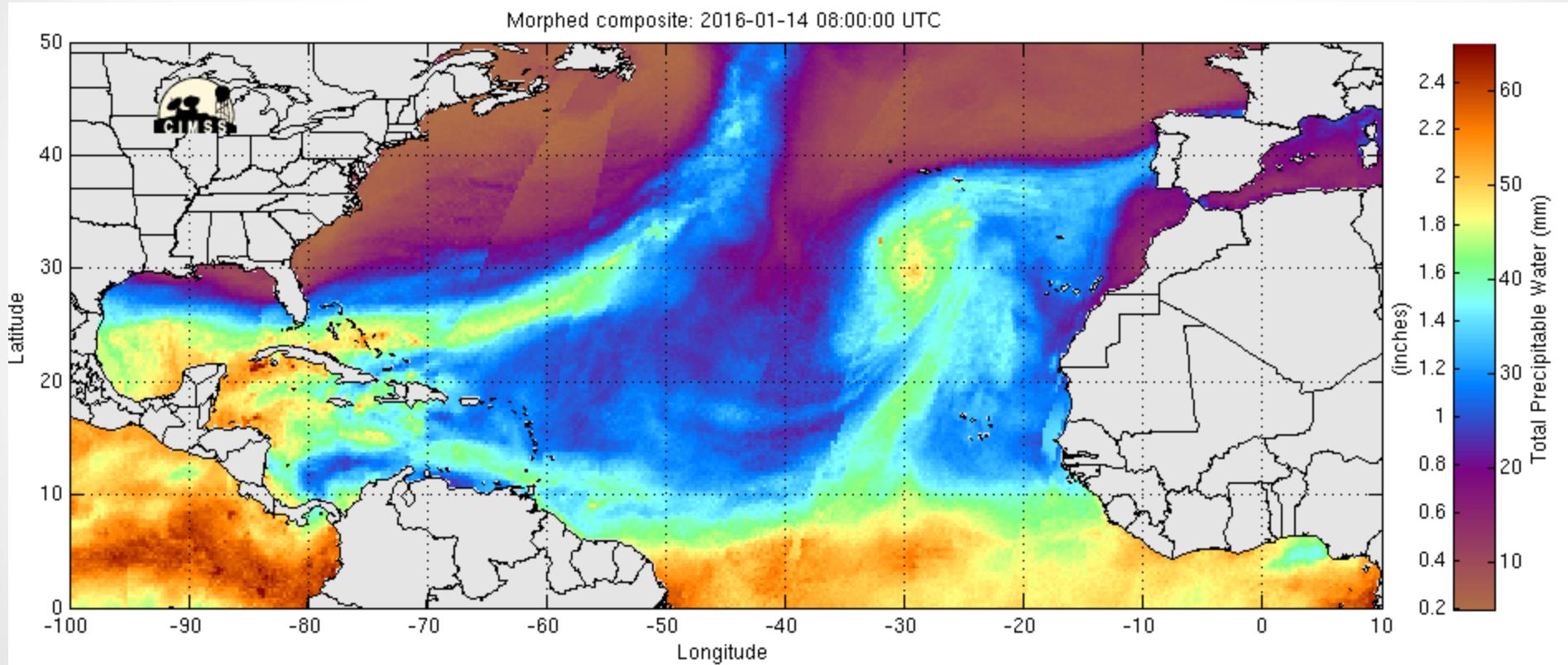


WHY IT RAINS

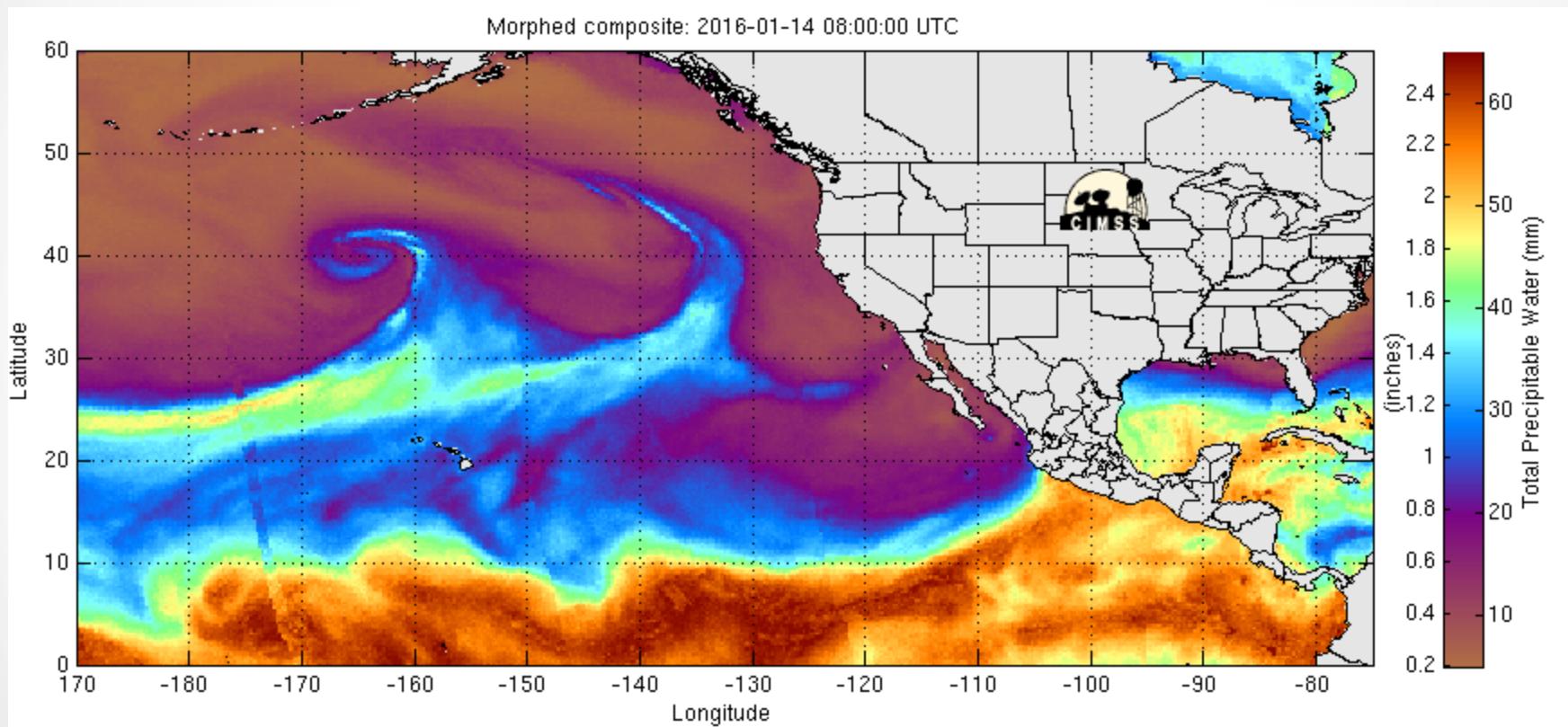


[NOAA/ESRL/PSD](#)
[CW3E SCRIPPS UCSD](#)

EAST COAST



WEST COAST

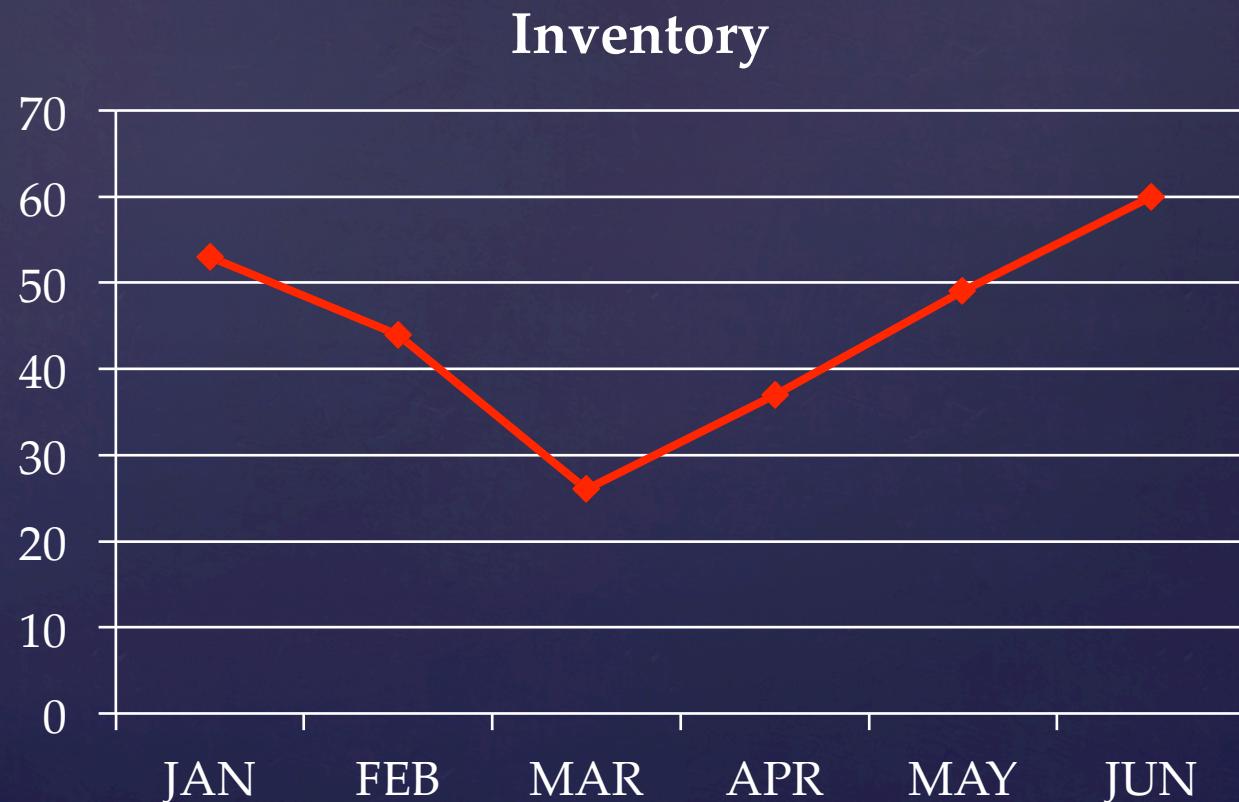


WHY DO WE NEED GRAPHICS?

{ Really necessary?

Which is better?

Month	JAN	FEB	MAR	APR	MAY	JUN
Inventory	53	44	26	37	49	60



WHY GRAPHS

“A picture is worth 1000 words”

- ¶ Discover or answer questions
- ¶ Make massive accessible & See data in context
- ¶ Pattern recognition, signal detection
- ¶ Support calculation, analysis and modeling
- ¶ Provide evidence or present argument
- ¶ Reasoning & Decision – making

and more...

THE ROLES OF GRAPHS

THINK LIKE A
DATA SCIENTIST

What it takes? {

GRAPHS ARE COMPARISON...

{ To find things unexpected

NEXT AMERICA

Percent of U.S. Population by Age Group, 1950-2060

Baby Boomers

MALE

FEMALE

1950

85+

75-79

50-54

25-29

0

4%

2%

0%

2%

4%

PEW RESEARCH CENTER

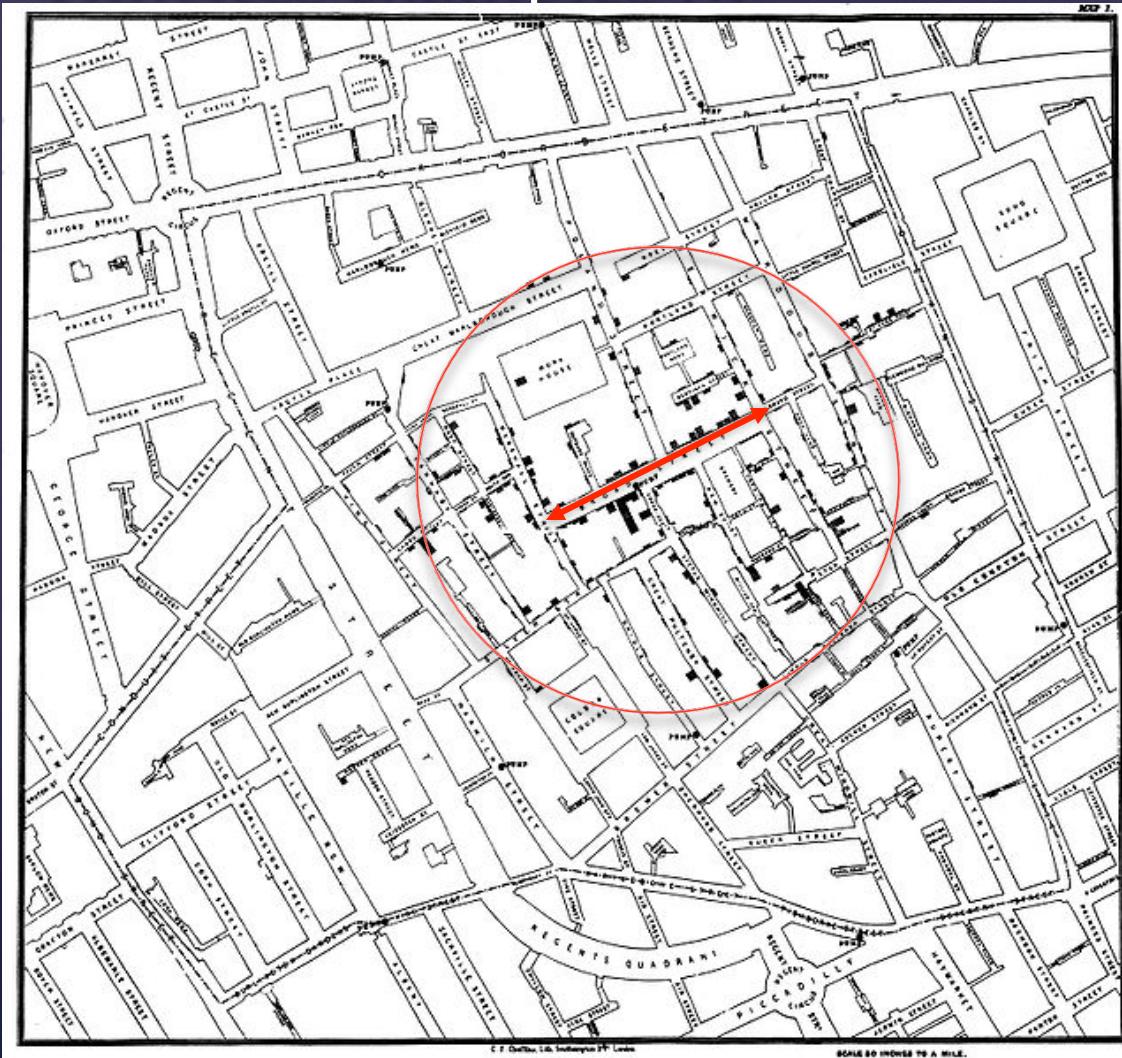
THE JOY OF STATS

BBC FOUR



GRAPHS ARE COMPARISONS

Timeless examples ...



John Snow's Dot Distribution Map of Broad Street Cholera Cases

source: [Wikimedia Commons](#)



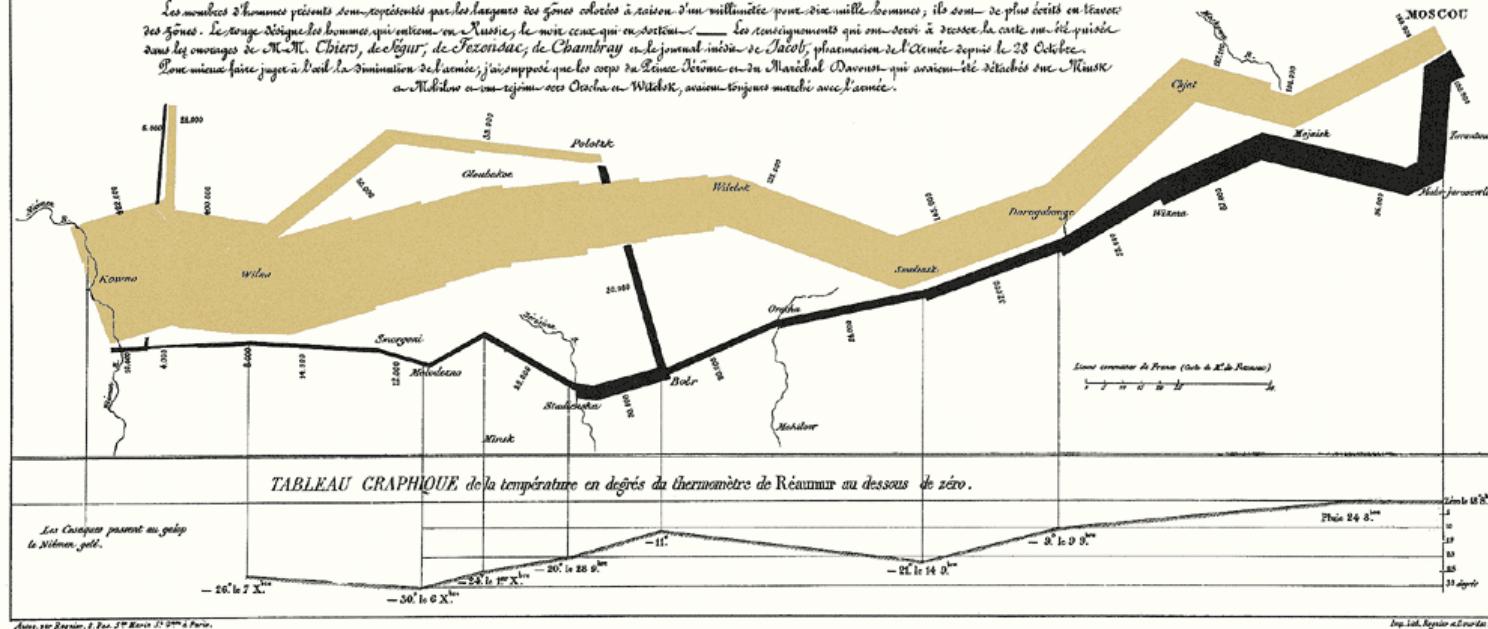
The Night Bivouac of Napoleon's Army during retreat from Russia in 1812. Oil on canvas. *Historical Museum, Moscow, Russia.*

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

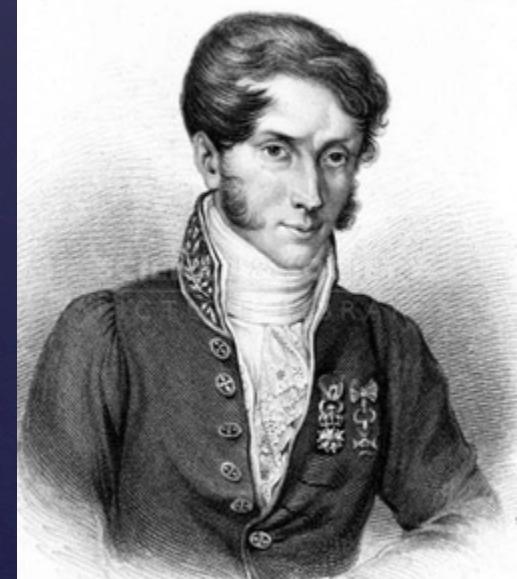
Deçue par M. Minard, Inspecteur Général resté sous le Charrois au siège de Paris, le 20 Novembre 1869

Les nombres d'hommes perdus sont représentés par les largures des zones colorées à raison d'un millimètre pour dix-mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui restent en Russie, le noir ceux qui rentrent. — Les renseignements qui me serviront à dresser la carte me dépendent des ouvrages de M. M. Chabot, de Liger, de Tocroisac, de Chambray, du journal intitulé de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

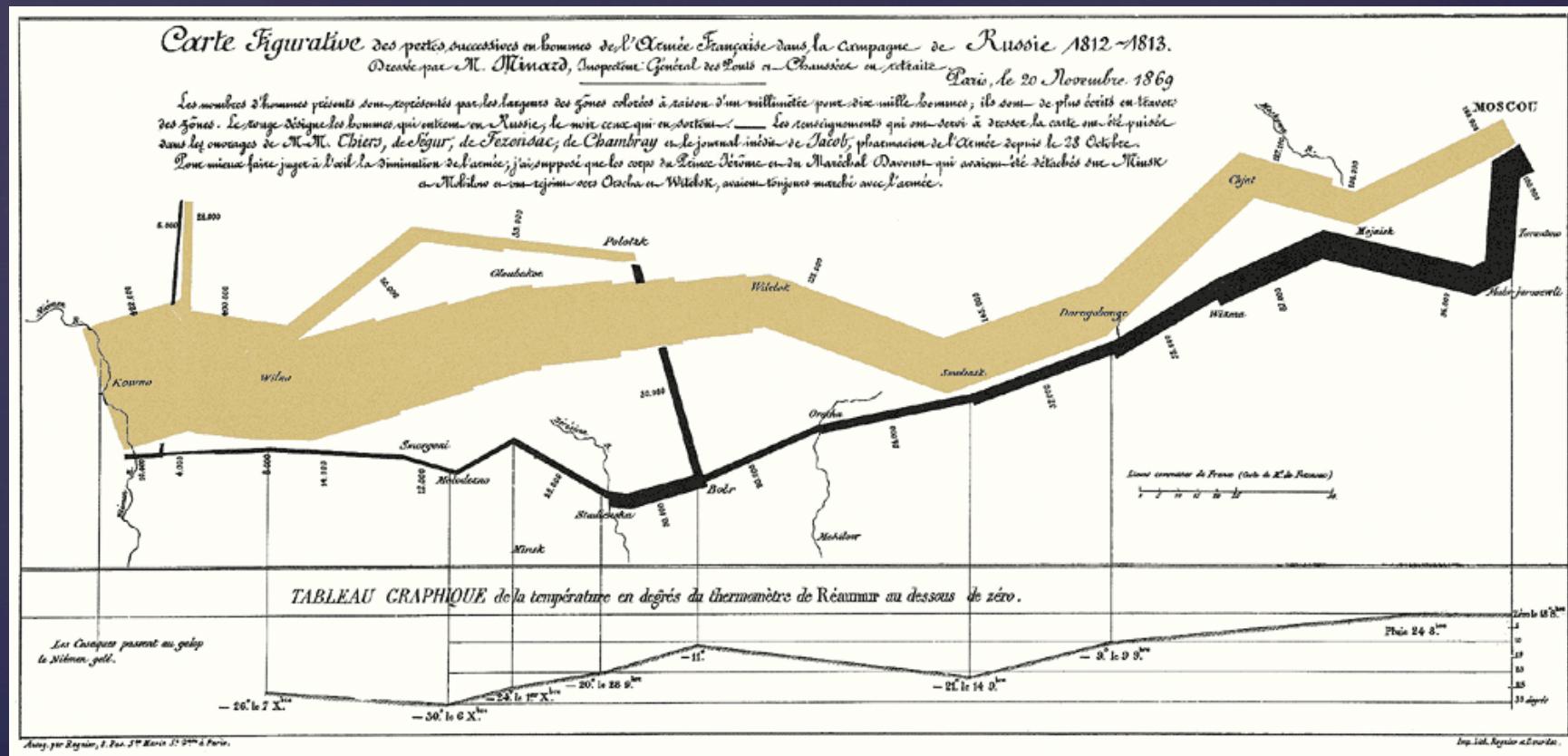
Pour mieux faire juger à l'échelle la diminution de l'armée, j'inscris sur les coqs la force établie en Russie, et du Maréchal Davout qui venait d'être détaché sur Moscou au début d'un régime avec Orléans au Witebsk, auquel tout juste marcha avec l'armée.



Charles Joseph Minard



Charles Minard's Visualization of Napoleon's 1812 March

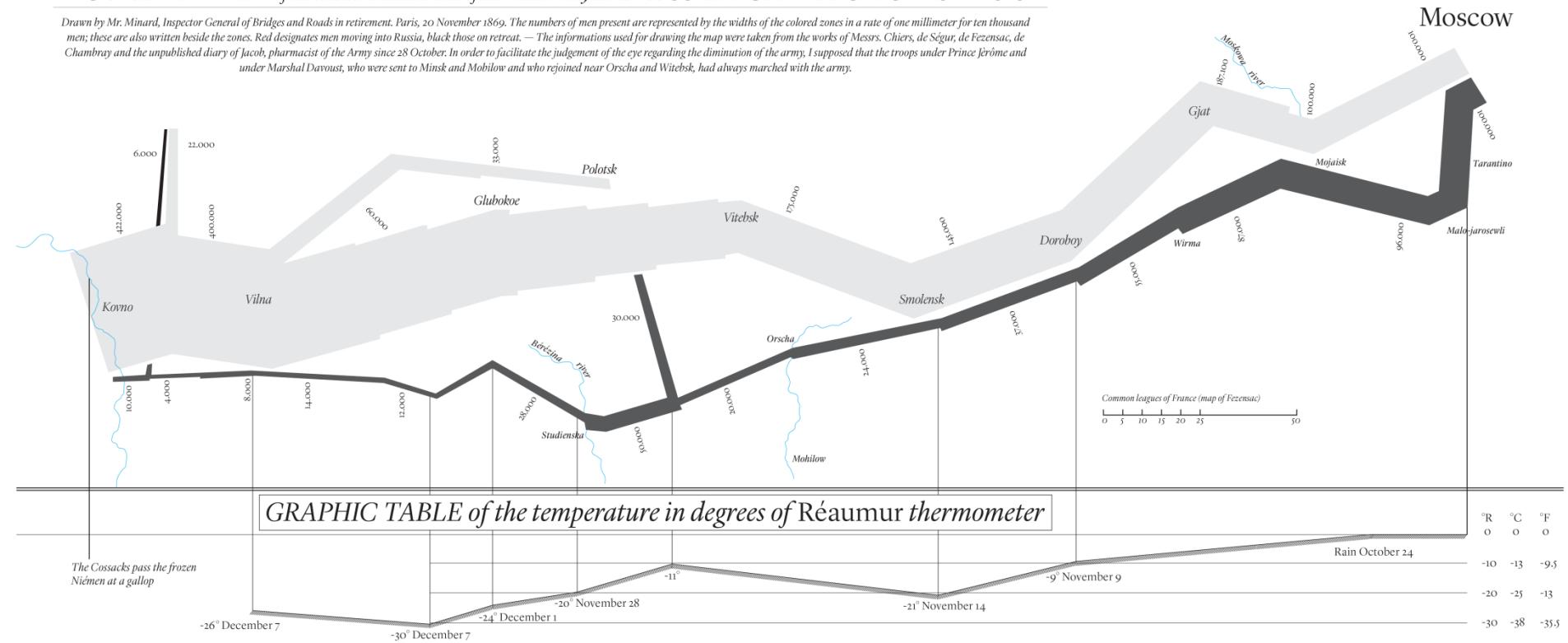


More can be found at: <http://www.edwardtufte.com/tufte/minard>

Modern redraw

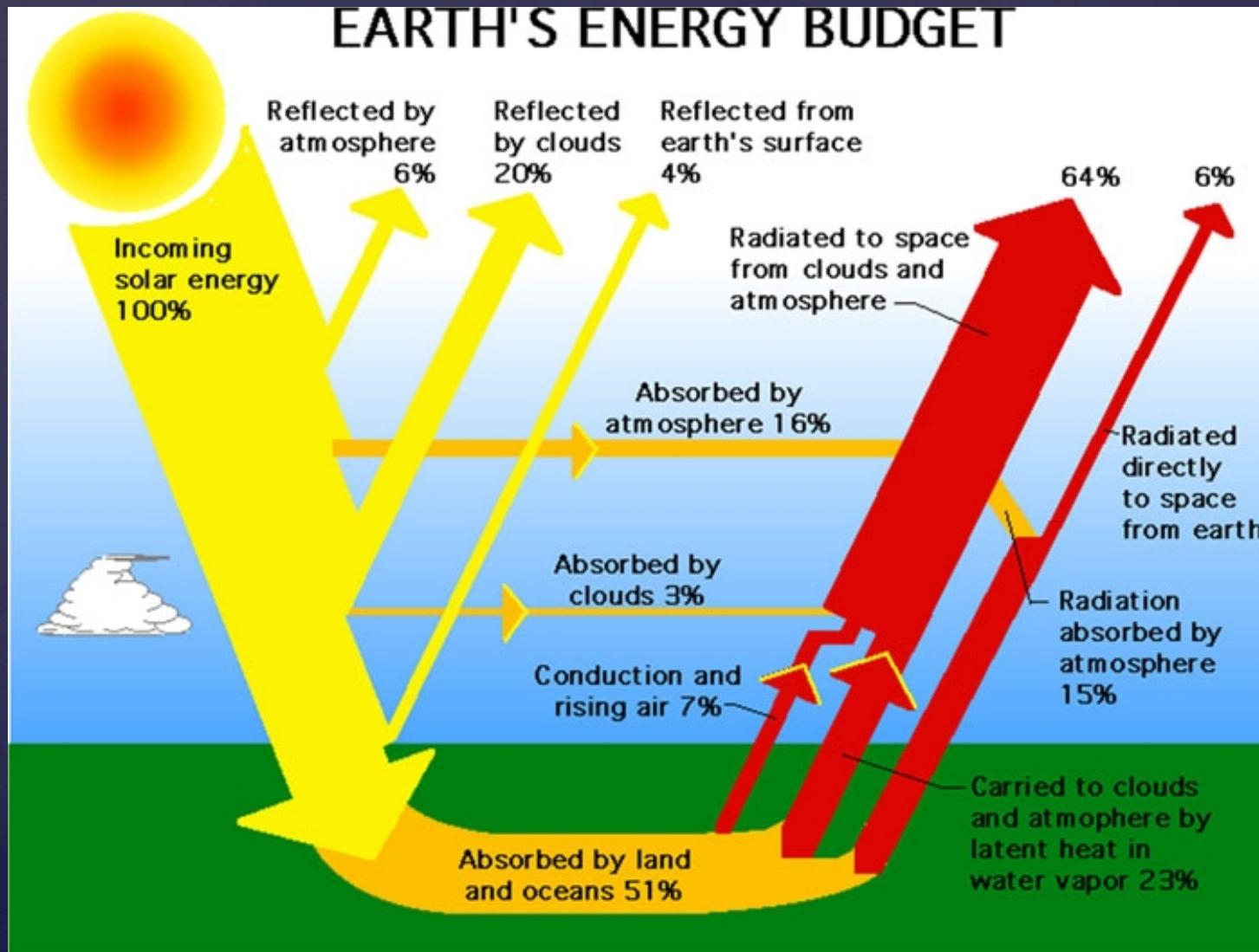
FIGURATIVE MAP of the successive losses in men of the French Army in the RUSSIAN CAMPAIGN OF 1812-1813

Drawn by Mr. Minard, Inspector General of Bridges and Roads in retirement. Paris, 20 November 1869. The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. — The informations used for drawing the map were taken from the works of Messrs. Chiers, de Ségar, de Fezensac, de Chambray and the unpublished diary of Jacob, pharmacist of the Army since 28 October. In order to facilitate the judgement of the eye regarding the diminution of the army, I supposed that the troops under Prince Jérôme and under Marshal Davout, who were sent to Minsk and Mobilow and who rejoined near Orscha and Witebsk, had always marched with the army.



“it may well be the best statistical graphic ever drawn”, Edward Tufte
 “Minard's map is a “gem” of information graphics, it is the World's Champion Graph”, Howard Wainer
 “His famous map was only one of 51 thematic maps he created during his lifetime, he was a true pioneer in many respects”, Arthur Robinson

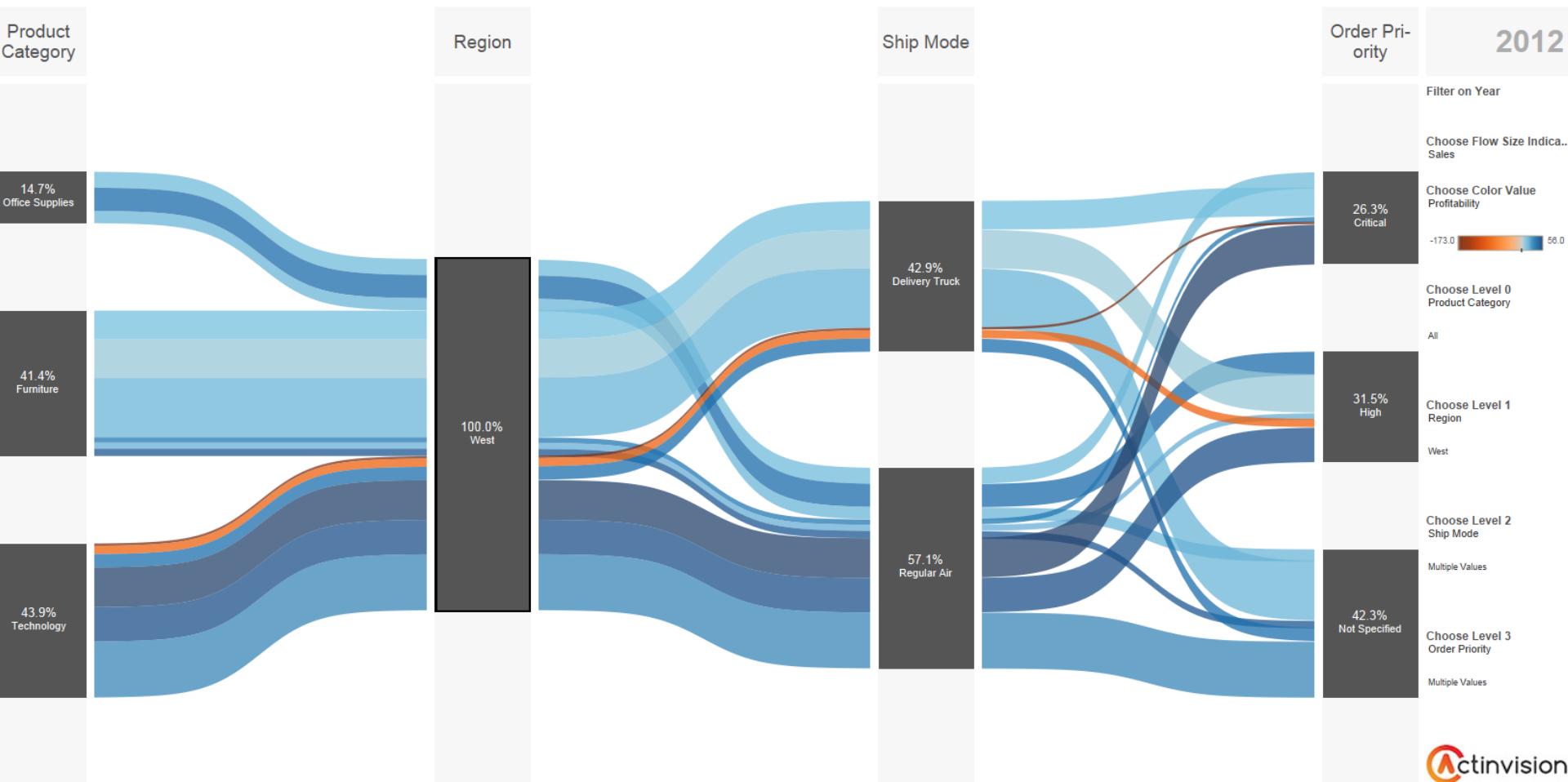
SANKEY DIAGRAM



SANKEY DIAGRAM



Superstore Interactive Sankey



SANKEY DIAGRAM

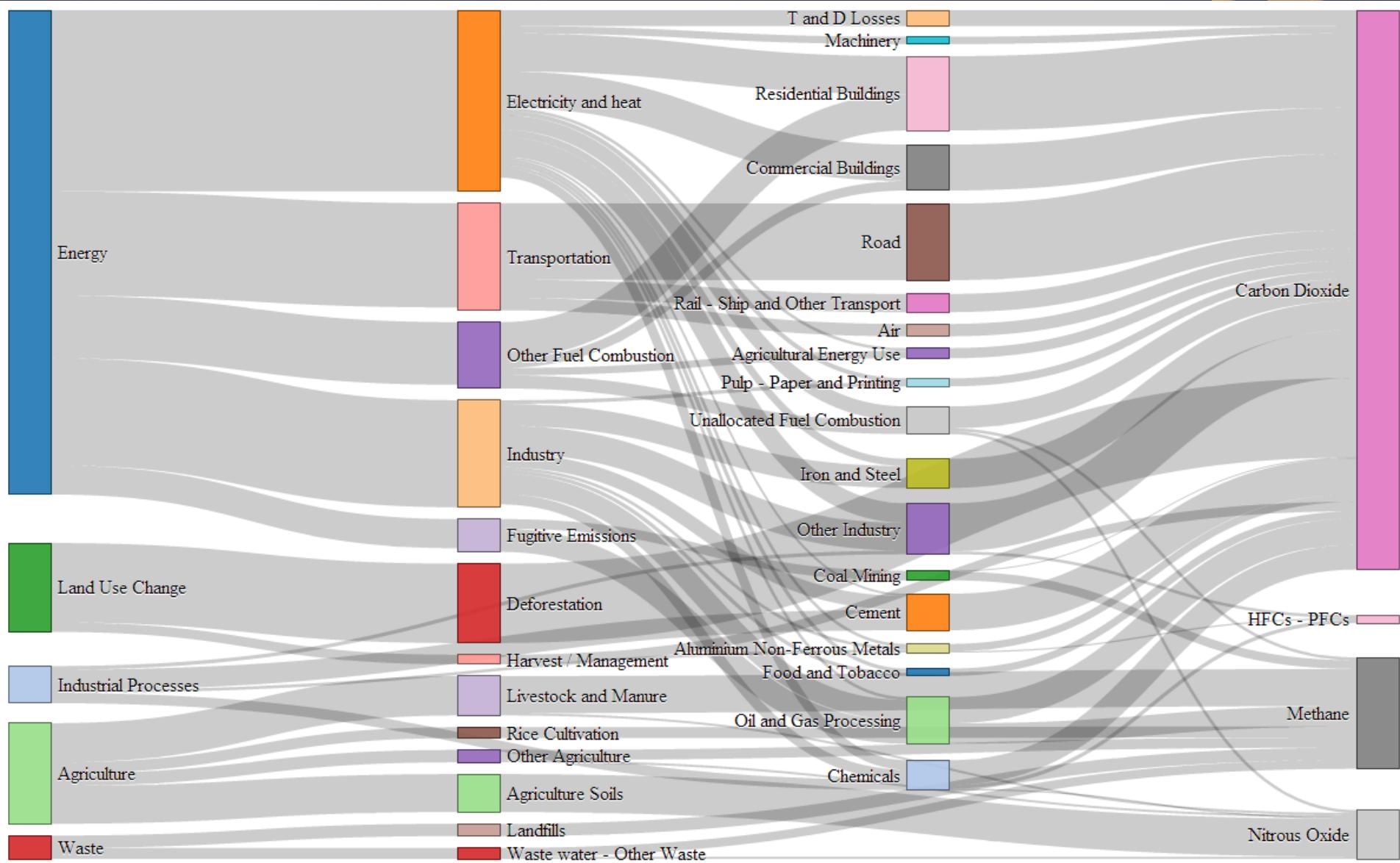
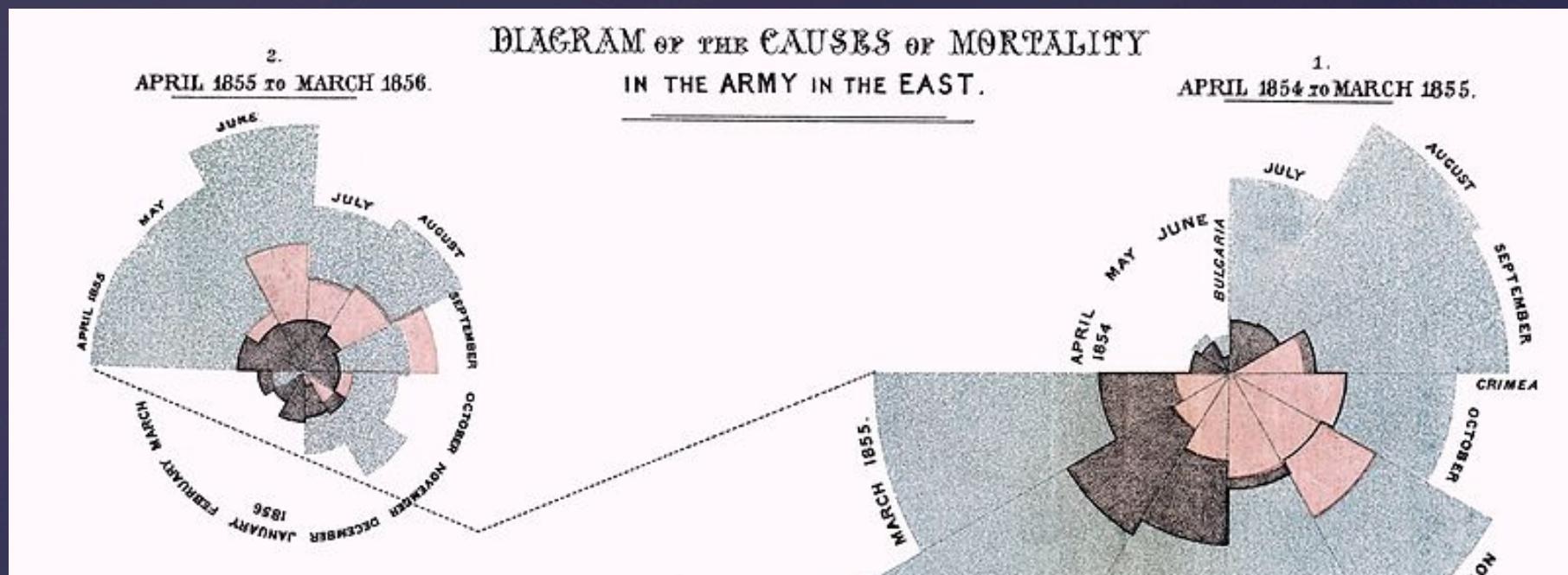


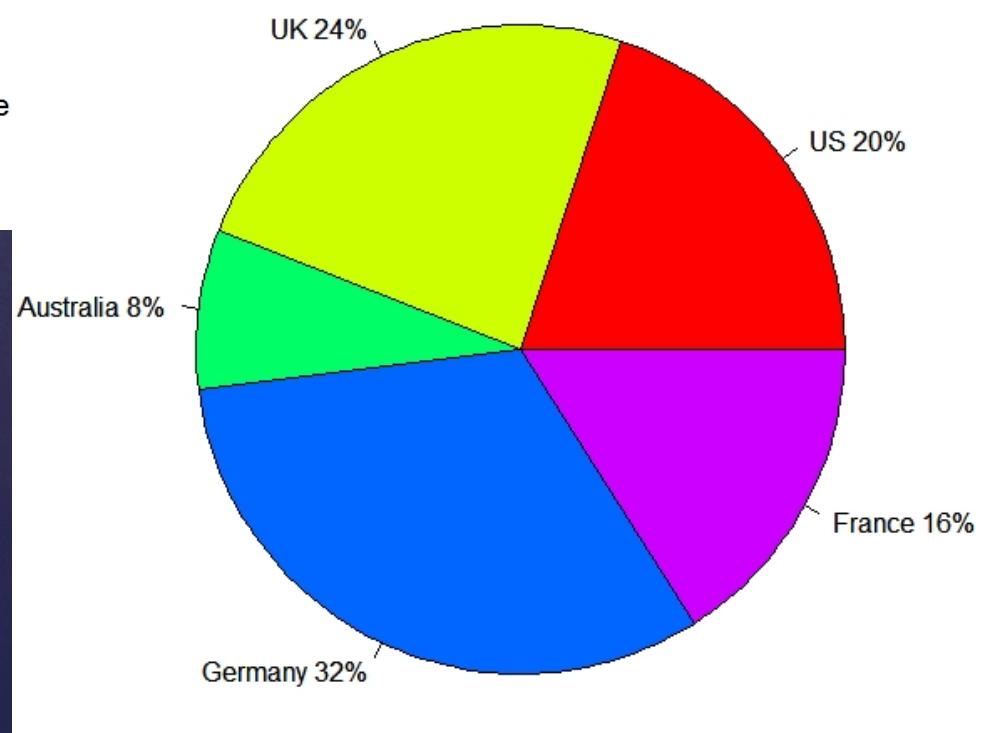
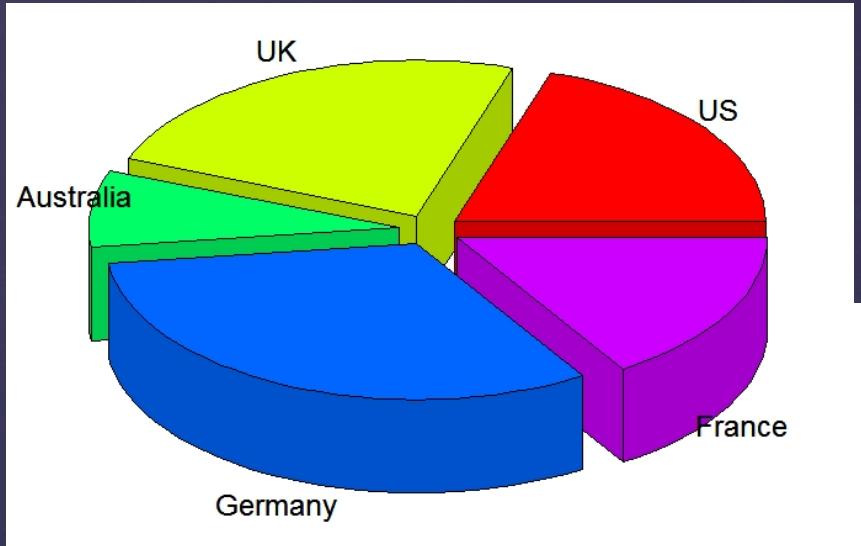
Diagram of the causes of mortality in the army in the East



The legend reads:

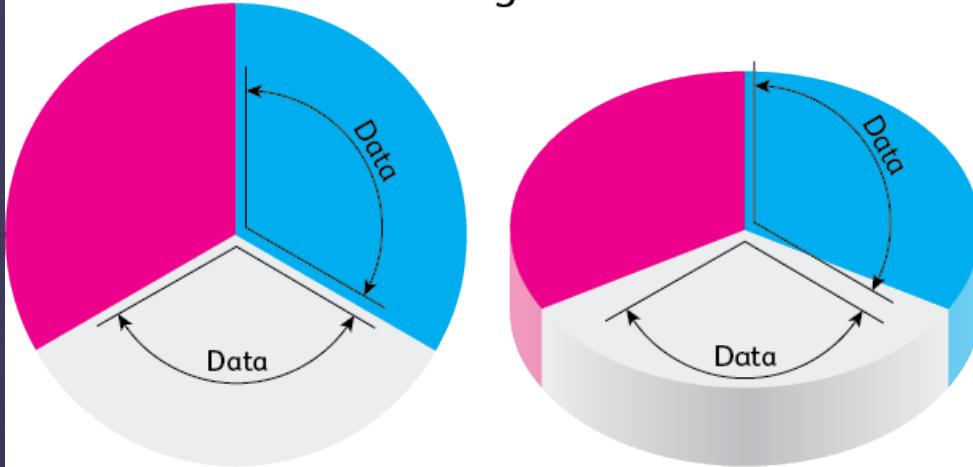
The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex. The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes. The black line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month. In October 1854, & April 1855, the black area coincides with the red, in January & February 1855, (*) the blue coincides with the black. The entire areas may be compared by following the blue, the red, & the black lines enclosing them.

There are caveats...



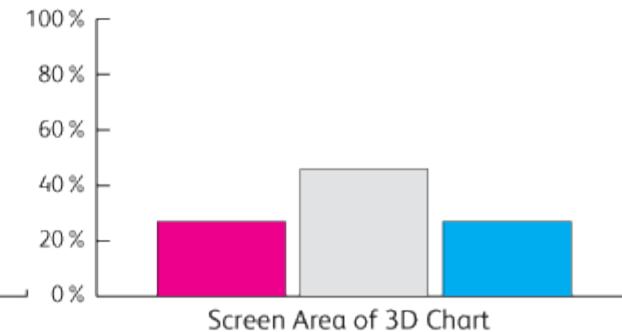
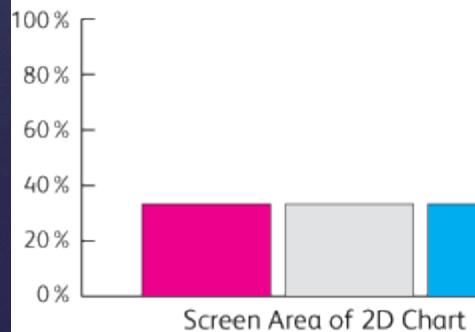
“2D’s Company, 3D’s a Crowd”

Angle

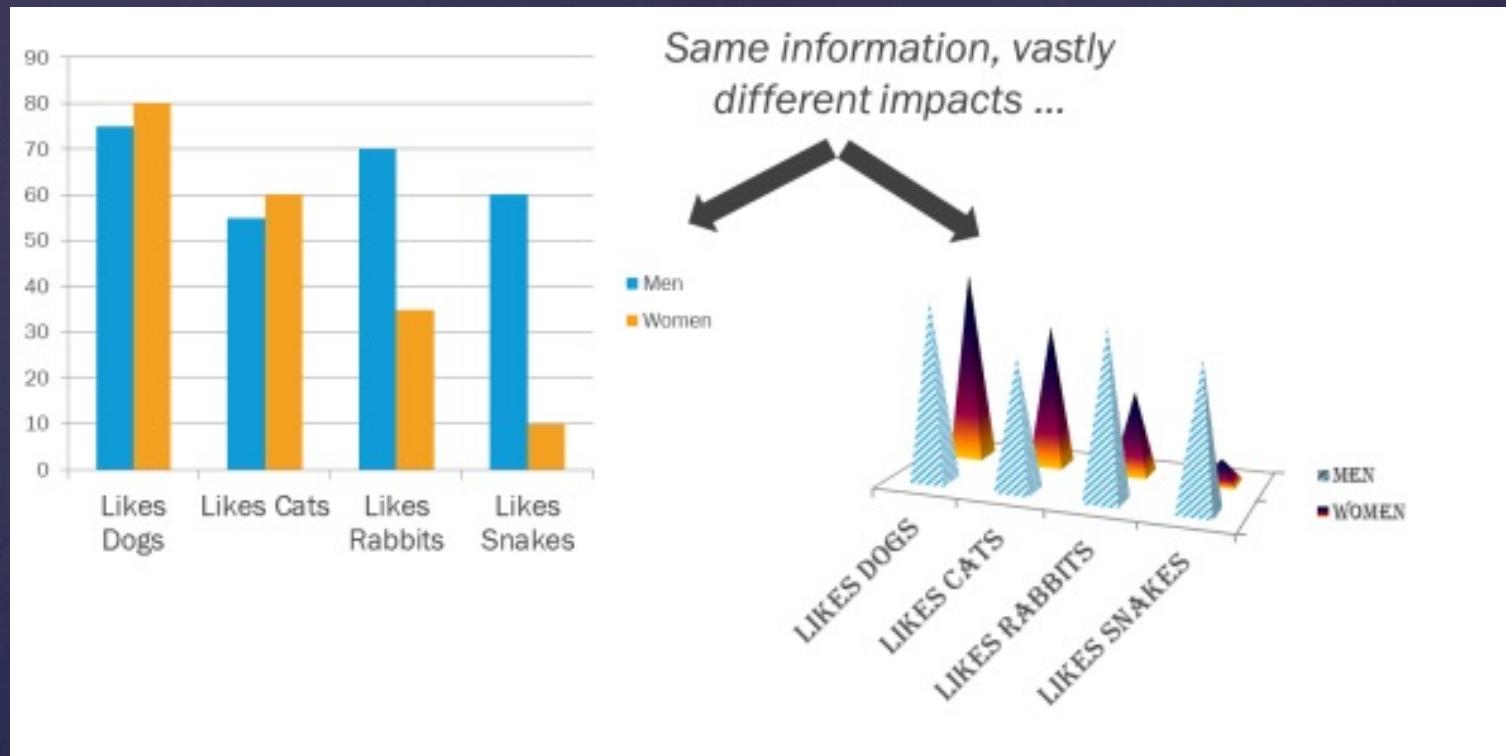


Violation of the
Area Principle

Area

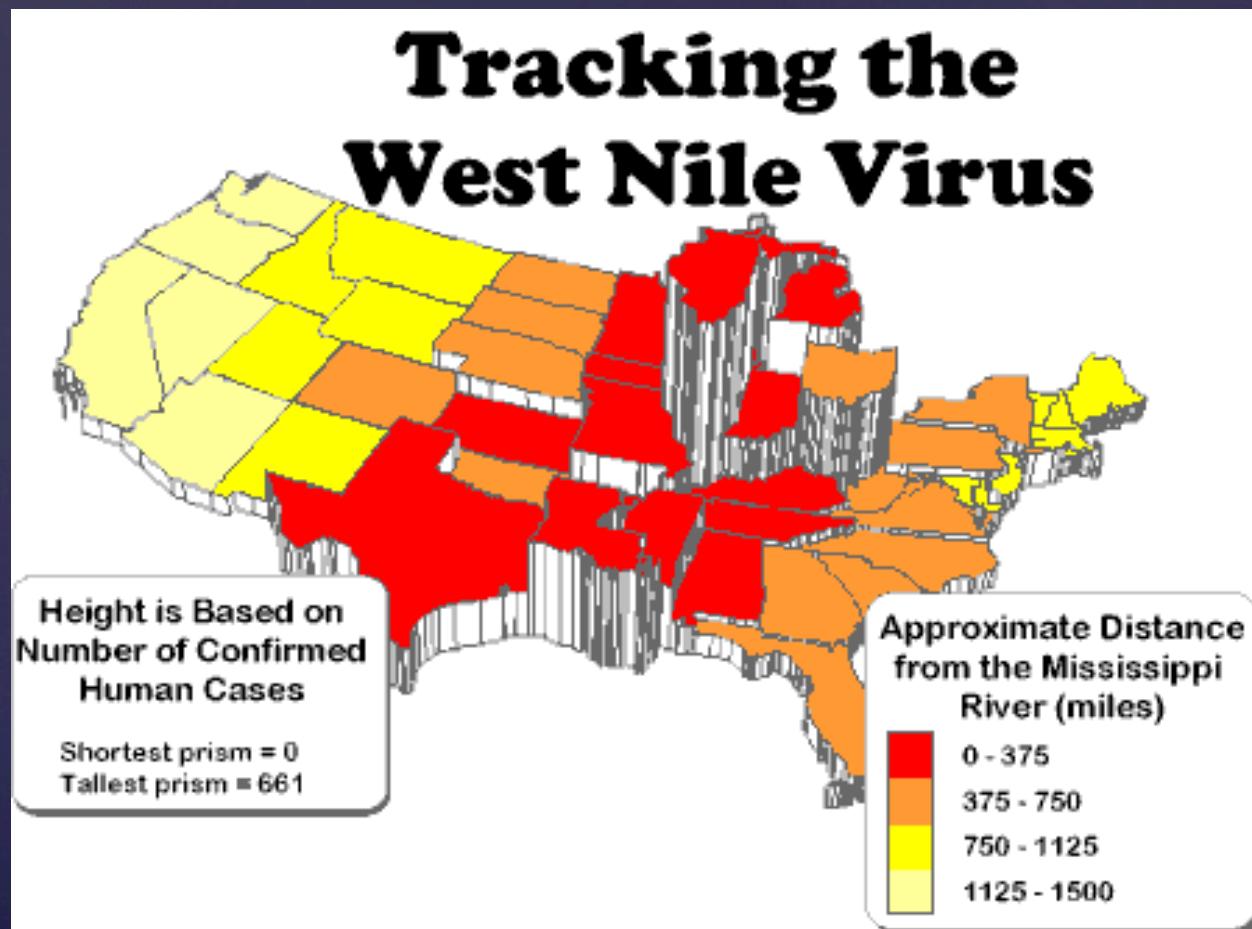


2D's Company, 3D's a Crowd – Pie chart is not alone



Source: <http://www.enterpriseappstoday.com>

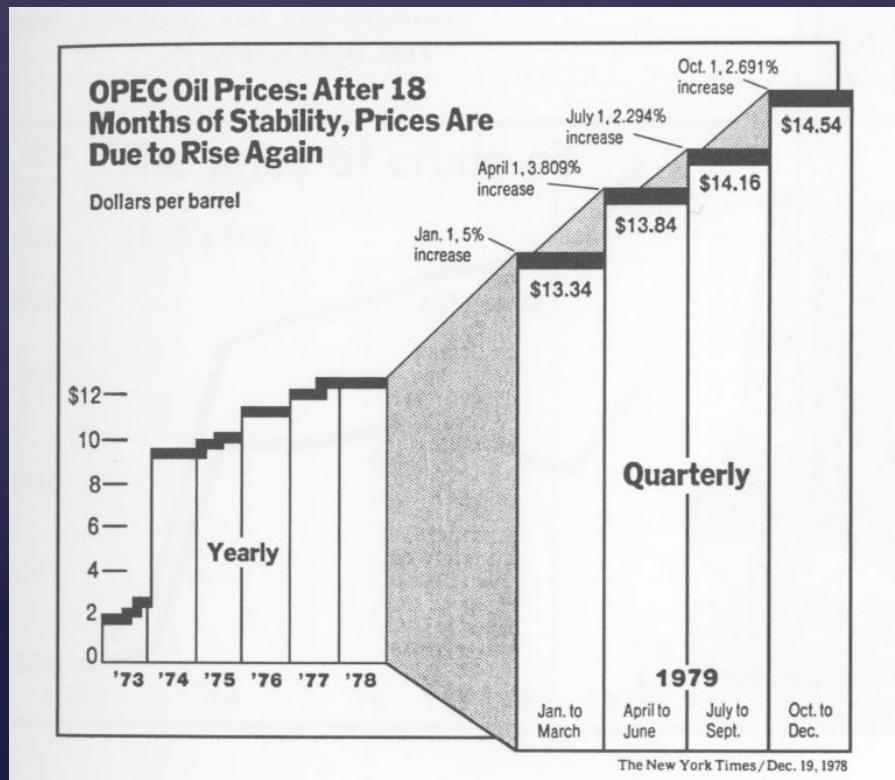
2D's Company, 3D's a Crowd



Four Principles

Ref: Edward Tufte, *The Visual Display of Quantitative Information*

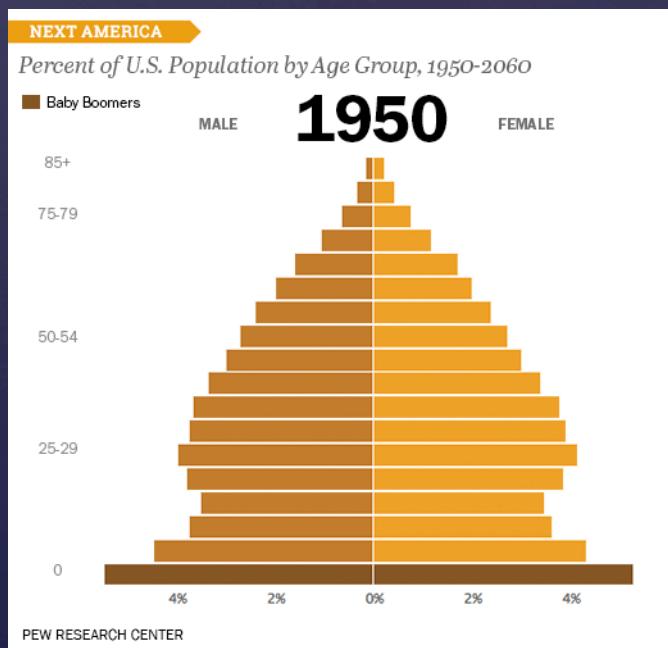
1. “Area Principle”
2. Labeling → Graphical distortion & ambiguity
3. Show data variation, not design variation



Four Principles

Ref: Edward Tufte, *The Visual Display of Quantitative Information*

1. “Area Principle” and beyond;
2. Labeling → Graphical distortion & ambiguity
3. Show data variation, not design variation
4. No. of Dimensions (Plots) ≤ No. of Variables (Data) ?



GitHub

{ Version control;
Code management;
Collaboration

1. Create a private repository if you just want your teammates see your codes before the submission (students' account automatically have this feature, open on with UNI@columbia.edu)
2. Send your teammates the link, they can all contribute to the repository

Collaboration

Get started with GitHub

Option 1:

1. Install git (
<https://git-scm.com/book/en/v2/Getting-Started-Installing-Git>)
2. Xcode – to initiate (you don't need this step, if your Xcode was initiated before.)
3. Open Terminal (ConEmu in Window)

Get started with GitHub

Option 2:

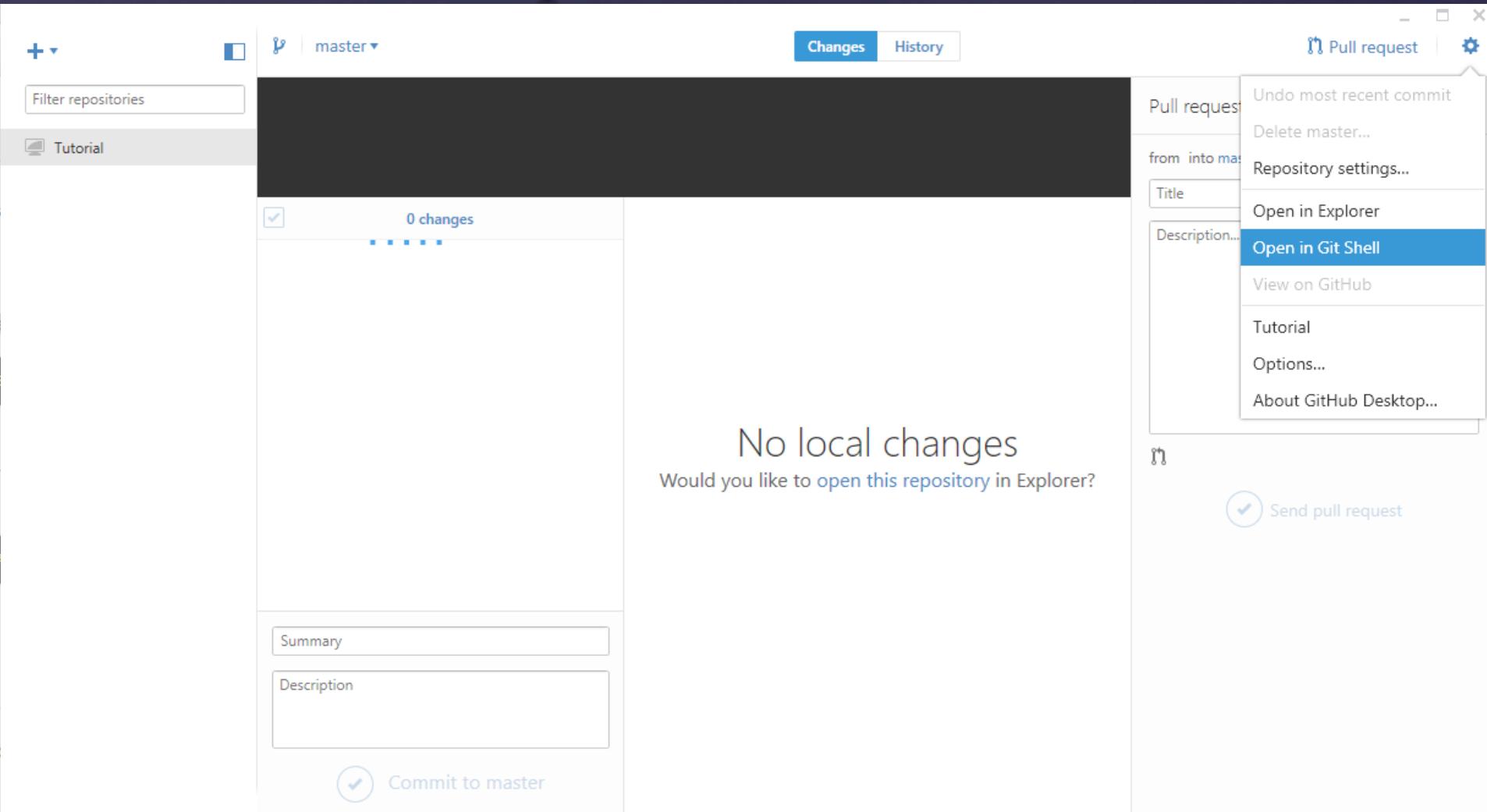
1. GitHub Desktop (
<https://guides.github.com/introduction/getting-your-project-on-github/>)
2. Download: (<https://desktop.github.com>)
Available for Mac and Windows
3. Install GitHub

Setting up Git

```
$ git config --global user.name "YOUR NAME"  
$ git config --global user.email "YOUR  
EMAIL ADDRESS"
```

Windows Users

Where to “Open in Git Shell” –



Existing Repository

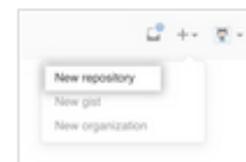
```
$ cd LOCAL  
$ git clone http://github.com/YOUR NAME/HW1  
$ git status # everything seems updated  
$ vi README.md  
$ git status # Now I have a new “README.md”  
$ git add README.md  
$ git commit -m”add README”  
$ git push
```

1. Has to do with version control;
2. If two work with the same folder, you both push your updates to the Github.com from version last updated at time A, one guy successfully push updates, the other will not.
3. To fix it: **You can only push updates with the old version is the same as on Github.com**

Just one problem ...

Create a new repository on GitHub

- 1 In the upper-right corner of any page, click **+**, and then click **New repository**.



- 2 Create a short, memorable name for your repository. For example, "hello-world".



- 3 Optionally, add a description of your repository. For example, "My first repository on GitHub."



- 4 Choose between creating a *public* or *private* repository.

- › **Public** repositories are a great choice for getting started! They're visible to any user on GitHub, so you can benefit from a collaborative community.
- › **Private** repositories require a little more setup. They're only available to you, the *repository owner*, as well as any *collaborators* you choose to share with. Private repositories are only available for paid accounts. For more information, see "[What plan should I choose?](#)"



- 5 Select **Initialize this repository with a README**.



- 6 Click **Create repository**.



Congratulations! You've successfully created your first repository, and initialized it with a *README* file.

Fork A Repository

- A *fork* is a copy of a repository.
- Forking a repository allows you to freely experiment with changes without affecting the original project.
- Forks are used to
 1. Propose changes to someone else's project
 2. Use someone else's project, **you can't change!**
- Good for debug:
 1. Fork the repository
 2. Make the fix
 3. Submit a *pull* request to the owner

How to Fork

- On GitHub

This screenshot shows a GitHub repository page for 'MRandomMax / EDAV'. The top navigation bar includes links for 'Pull requests', 'Issues', and 'Gist'. Below the header, there's a search bar and a user profile icon. The main content area displays the repository name 'MRandomMax / EDAV' and a 'Code' tab selected. Other tabs include 'Issues (0)', 'Pull requests (0)', 'Wiki', 'Pulse', 'Graphs', and 'Settings'. A red box highlights the 'Fork' button, which has 13 forks. Below the tabs, repository statistics are shown: '2 commits', '1 branch', '0 releases', and '1 contributor'. A 'New pull request' button is also visible. The repository details section shows files like 'Lecture Notes' (version 'lec1', 3 days ago) and 'README.md' (version 'I', 2 days ago). A note at the bottom encourages adding a README, with a 'Add a README' button.

MRandomMax / EDAV

Code Issues (0) Pull requests (0) Wiki Pulse Graphs Settings

2 commits 1 branch 0 releases 1 contributor

New pull request New file Find file HTTPS https://github.com/MRando

MRandomMax | Latest commit c337496 2 days ago

Lecture Notes lec1 3 days ago

README.md I 2 days ago

Add a README

Keep your fork synced

- In Terminal

```
$ git clone https://github.com/USERNAME/EDAV
```

Now you have a local copy of your fork of my EDAV repository.

Then you need to configure Git to sync your fork with my original EDAV repository

Configure Git to sync your fork with the original repository

```
$ cd LOCAL
$ git remote -v
origin https://github.com/YOUR_USERNAME/EDAV.git (fetch)
origin https://github.com/YOUR_USERNAME/EDAV.git (push)
$ git remote add upstream https://github.com/MRandomMax/EDAV.git
$ git remote -v
origin https://github.com/YOUR_USERNAME/EDAV.git (fetch)
origin https://github.com/YOUR_USERNAME/EDAV.git (push)
origin https://github.com/MRandomMax/EDAV.git (fetch)
origin https://github.com/MRandomMax/EDAV.git (push)
```

Syncing a fork

```
$ git fetch upstream
remote: Counting objects: 75, done.
remote: Compressing objects: 100% (53/53), done.
remote: Total 62 (delta 27), reused 44 (delta 9)
Unpacking objects: 100% (62/62), done.
From https://github.com/ORIGINAL_OWNER/ORIGINAL_REPOSITORY
 * [new branch]      master      -> upstream/master

$ git checkout master
Switched to branch 'master'

$ git merge upstream/master
Updating a422352..5fdff0f
Fast-forward
 README                      |    9 -----
 README.md                    |    7 ++++++
 2 files changed, 7 insertions(+), 9 deletions(-)
 delete mode 100644 README
 create mode 100644 README.md

$ git merge upstream/master
Updating 34e91da..16c56ad
Fast-forward
 README.md                    |    5 +++--
 1 file changed, 3 insertions(+), 2 deletions(-)
```