

# Data Visualization with R

@MatthewRenze

#KCDC

*Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.*  
 Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite à Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séjourné, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk en Mobilow et se rejoignirent vers Orscha et Witebsk, avaient toujours marché avec l'armée.

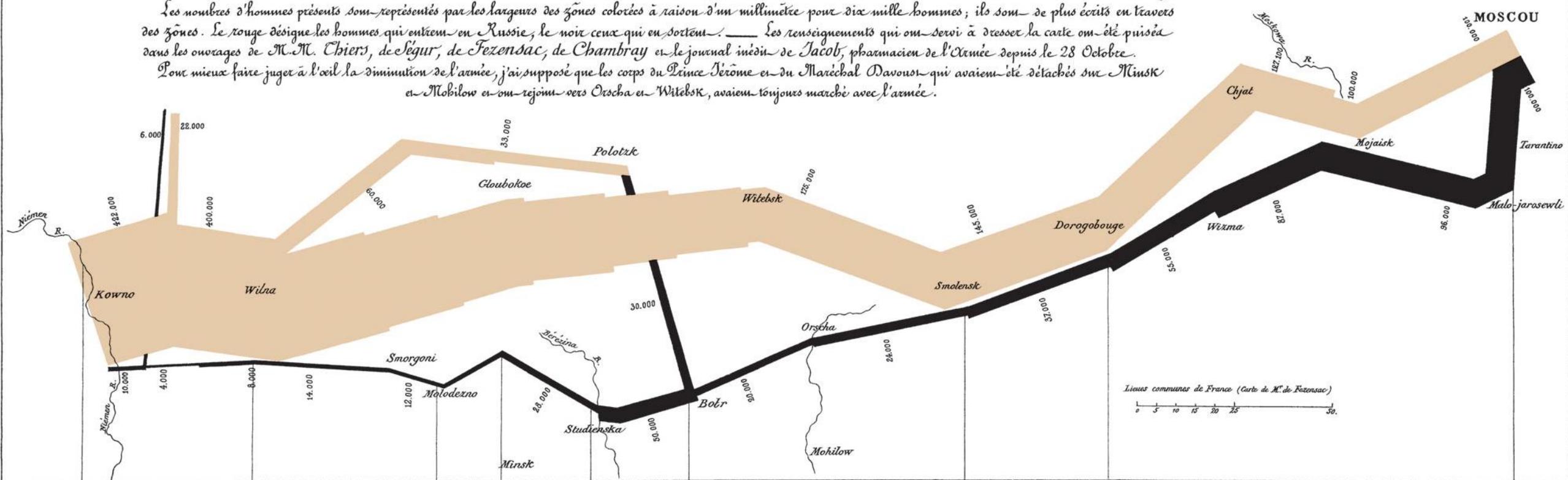


TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les cosaques passent au galop  
le Niémen gelé.

— 26° le 7 X<sup>bre</sup>.  
 — 30° le 6 X<sup>bre</sup>.  
 — 24° le 1<sup>er</sup> X<sup>bre</sup>.  
 — 20° le 28 9<sup>bre</sup>.  
 — 11°.

— 21° le 14 9<sup>bre</sup>.  
 — 8° le 9 9<sup>bre</sup>.

Zéro le 18 8<sup>bre</sup>.  
 Pluie 24 8<sup>bre</sup>.  
 5  
 10  
 15  
 20  
 25  
 30 degrés

*Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.*  
 Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite à Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séjourné, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mogilow et se rejoignaient vers Orscha et Witebsk, avaient toujours marché avec l'armée.

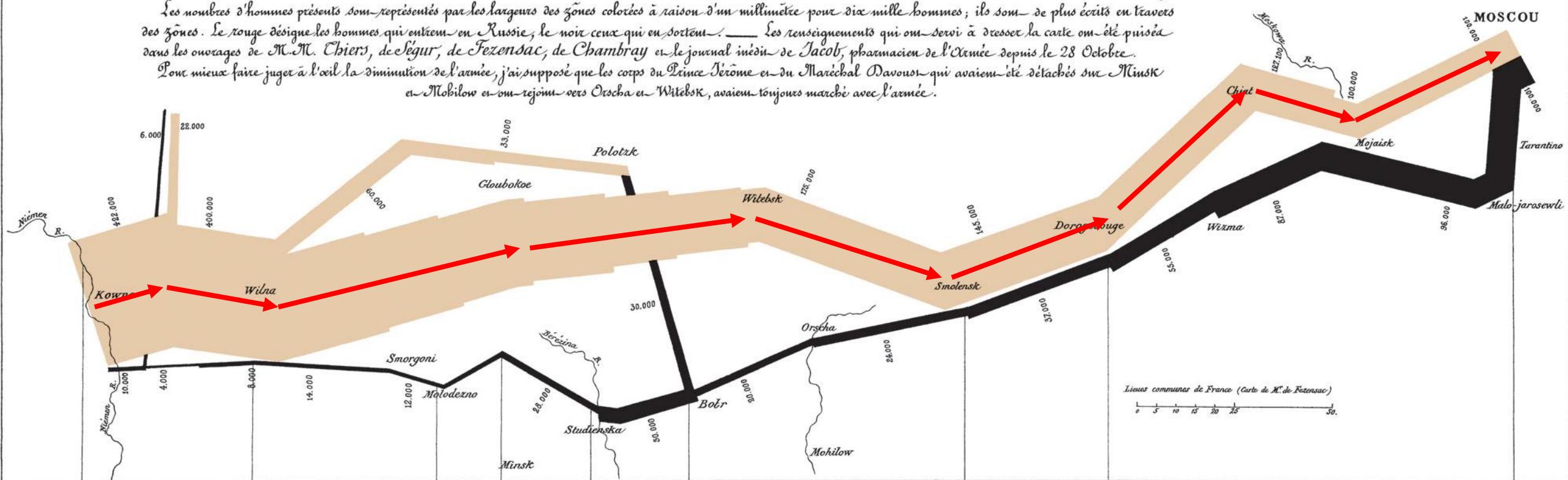


TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les cosaques passent au galop  
le Niemen gelé.

— 26° le 7 X<sup>bre</sup>.  
 — 30° le 6 X<sup>bre</sup>.  
 — 24° le 1<sup>er</sup> X<sup>bre</sup>.  
 — 20° le 28 9<sup>bre</sup>.  
 — 11°.

— 21° le 14 9<sup>bre</sup>.  
 — 8° le 9 9<sup>bre</sup>.

Zéro le 18 8<sup>bre</sup>.  
 Pluie 24 8<sup>bre</sup>.  
 5  
 10  
 15  
 20  
 25  
 30 degrés

*Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.*  
 Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite à Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séjourné, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk en Mobilow et se rejoignirent vers Orscha et Witebsk, avaient toujours marché avec l'armée.

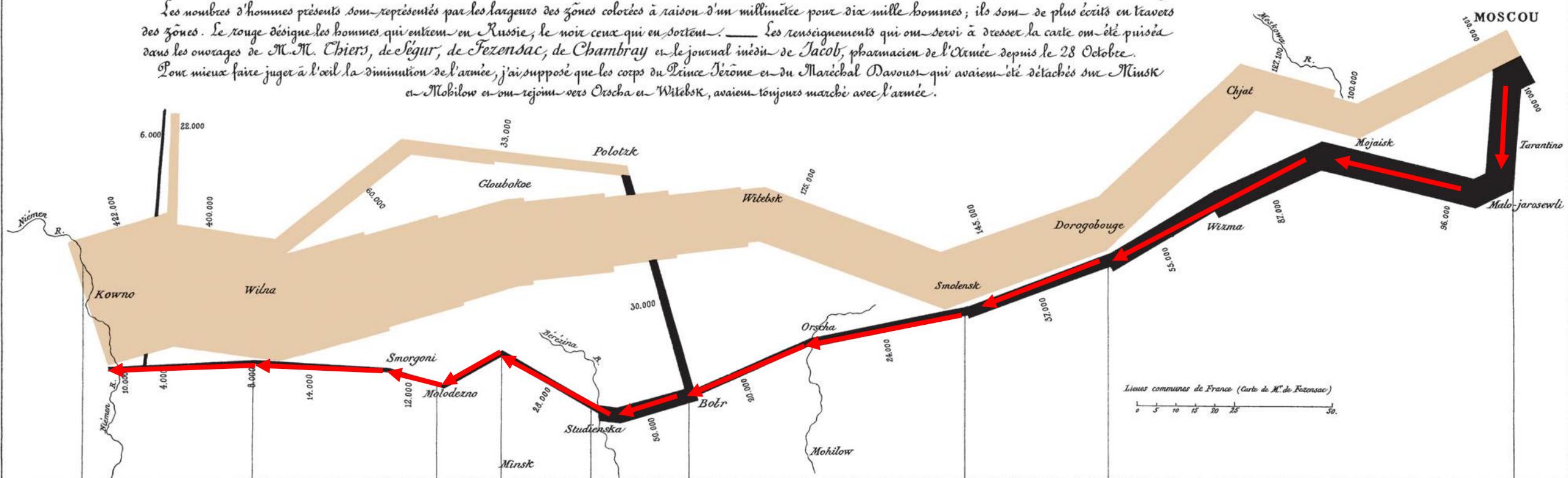


TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les cosaques passent au galop  
le Niemen gelé.

— 26° le 7 X<sup>bre</sup>.  
 — 30° le 6 X<sup>bre</sup>.  
 — 24° le 1<sup>er</sup> X<sup>bre</sup>.  
 — 20° le 28 9<sup>bre</sup>.  
 — 11°.

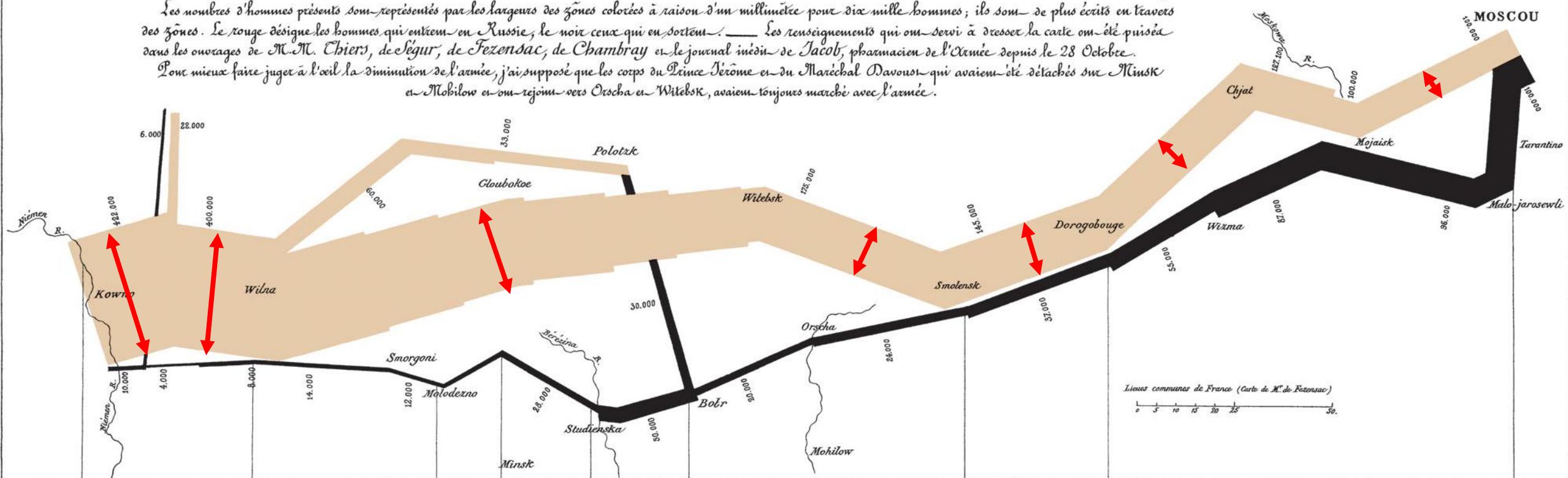
— 21° le 14 9<sup>bre</sup>.  
 — 8° le 9 9<sup>bre</sup>.

Zéro le 18 8<sup>bre</sup>.  
 Pluie 24 8<sup>bre</sup>.  
 5  
 10  
 15  
 20  
 25  
 30 degrés

*Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.*  
 Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite à Paris, le 20 Novembre 1869.

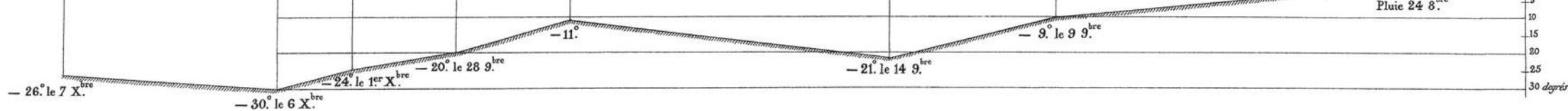
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séjourné, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mogilow et se rejoignirent vers Orscha et Witebsk, avaient toujours marché avec l'armée.



Lieux communs de France (Carte de M. de Fezensac)

TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

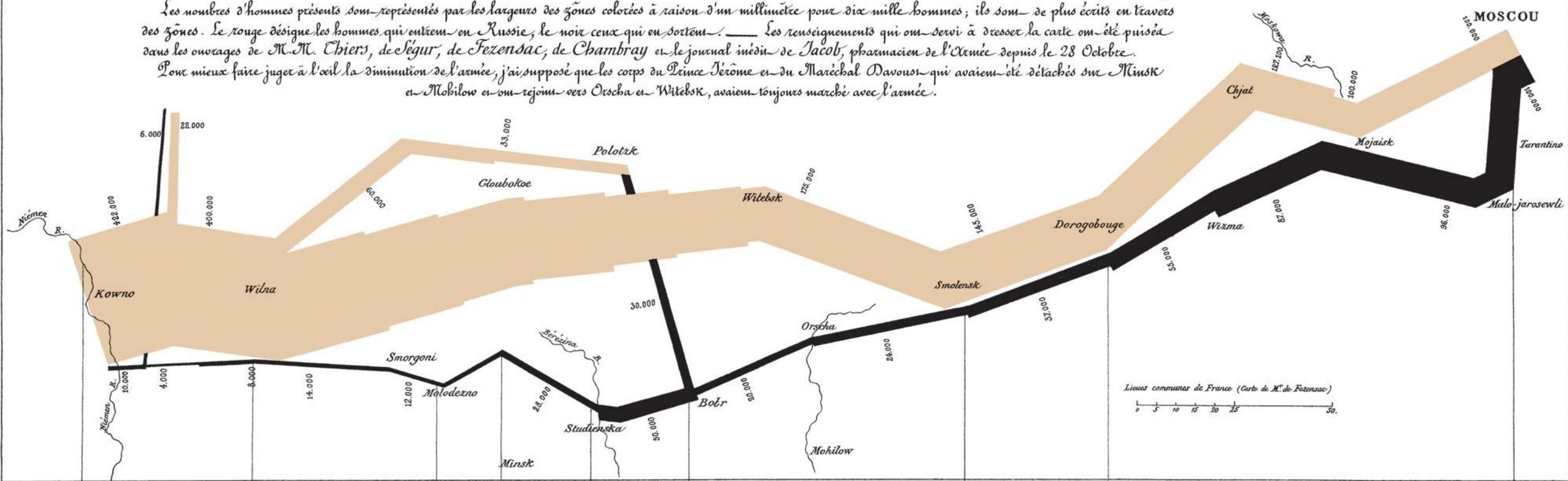


Les cosaques passent au galop  
le Niemen gelé.

*Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.*  
*Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite* Paris, le 20 Novembre 1869.

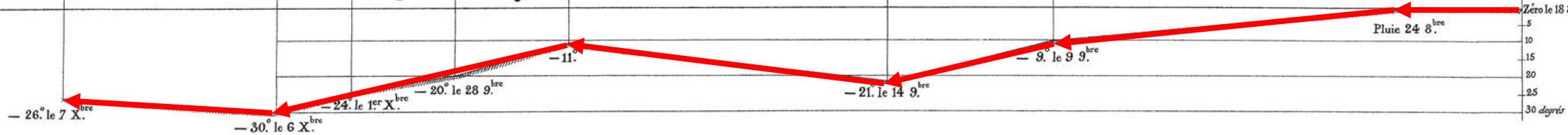
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Segur, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mobilow et ont rejoint vers Orsha en Witebsk, avaient toujours marché avec l'armée.



*TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.*

*Les Cosaques passent au galop  
le Niémen gelé.*



*Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.*  
 Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite à Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séjourné, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk en Mobilow et se rejoignirent vers Orscha et Witebsk, avaient toujours marché avec l'armée.

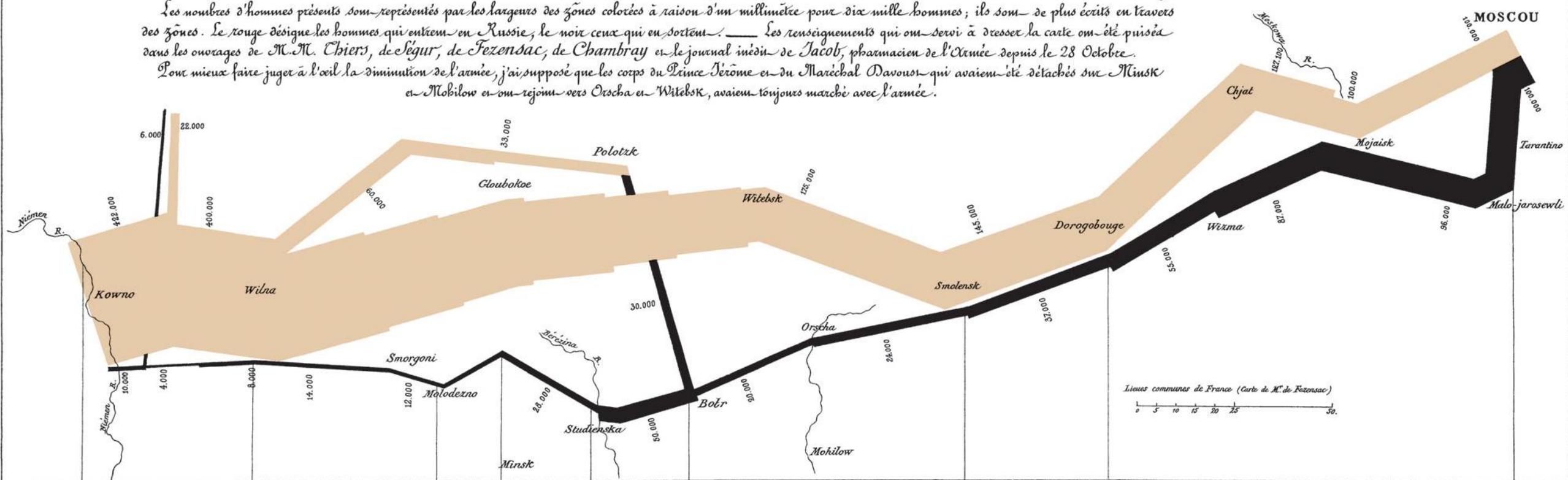


TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les cosaques passent au galop  
le Niemen gelé.

— 26° le 7 X<sup>bre</sup>.  
 — 30° le 6 X<sup>bre</sup>.  
 — 24° le 1<sup>er</sup> X<sup>bre</sup>.  
 — 20° le 28 9<sup>bre</sup>.  
 — 11°.

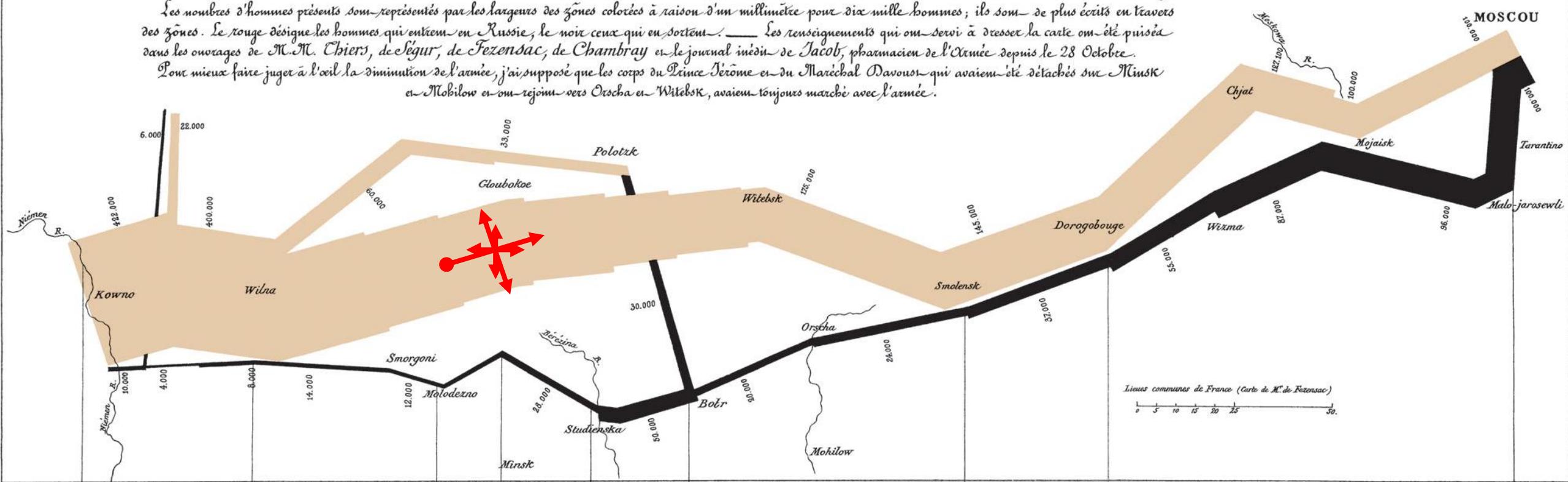
— 21° le 14 9<sup>bre</sup>.  
 — 8° le 9 9<sup>bre</sup>.

Zéro le 18 8<sup>bre</sup>.  
 Pluie 24 8<sup>bre</sup>.  
 5  
 10  
 15  
 20  
 25  
 30 degrés

*Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.*  
 Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite  
 Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séjourné, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

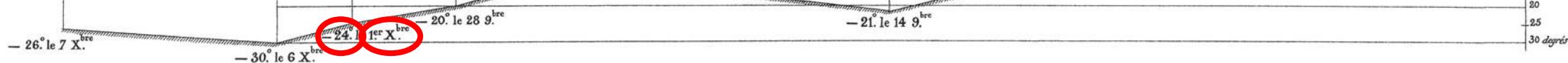
Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk en Mobilow et se rejoignaient vers Orscha et Witebsk, avaient toujours marché avec l'armée.

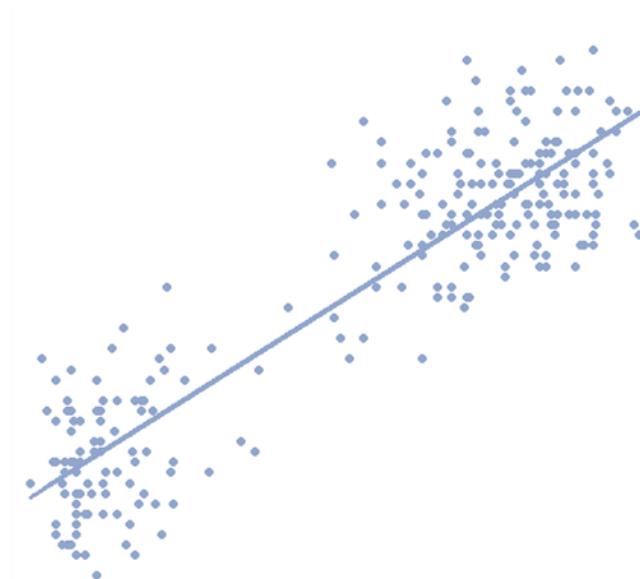
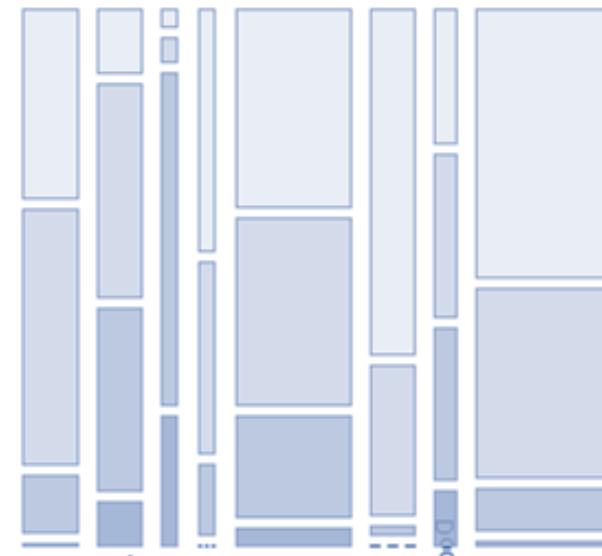
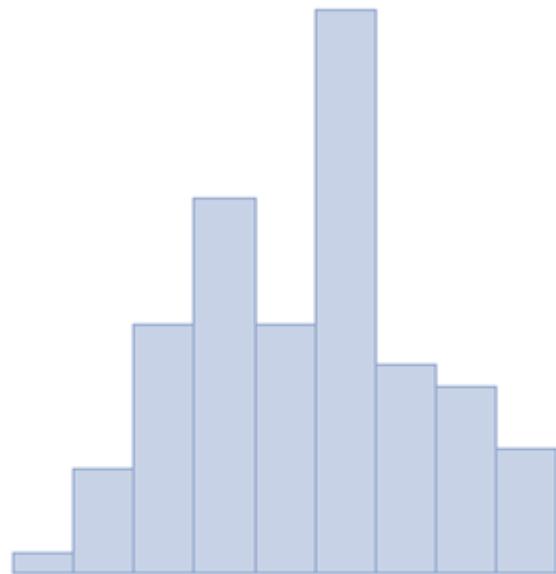


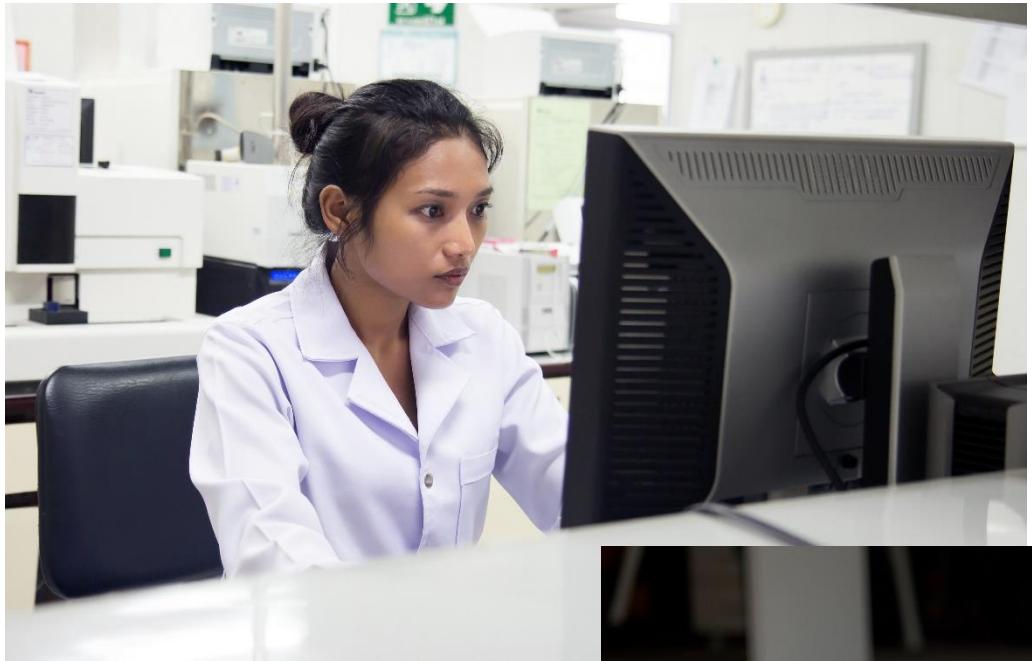
Lieux communs de France (Carte de M. de Fezensac)

TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les cosaques passent au galop  
le Niemen gelé.







# Overview

Introduction to R

Intro to Data Visualization

Types of Data Visualizations

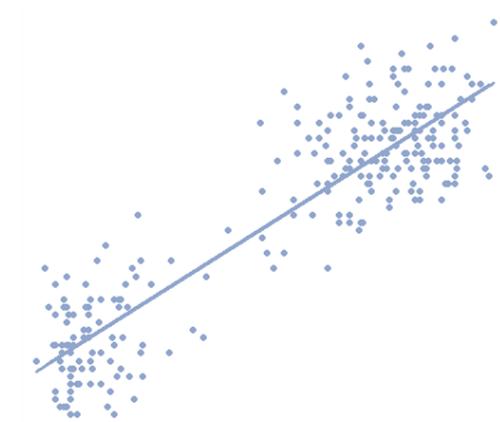
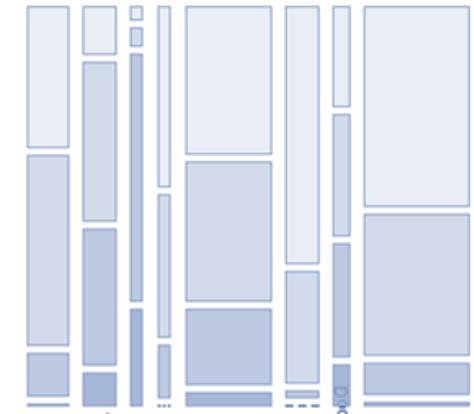
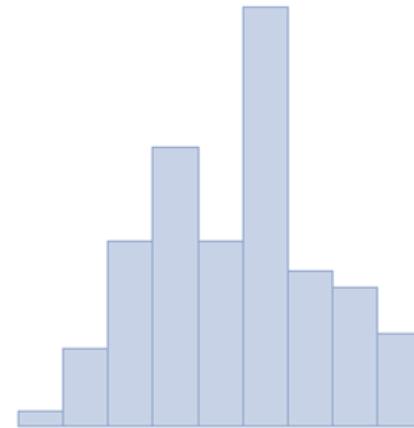
- Visualizing One Variable

- Visualizing Two Variables

- Visualizing Many Variables

- Visualizing Other Types of Data

Beyond the Basics



# About Me

Independent software consultant

Education

B.S. in Computer Science

B.A. in Philosophy

Community

Pluralsight Author

ASPIInsider

Public Speaker

Open-Source Software

IOWA STATE  
UNIVERSITY



## TITANIUM SPONSORS



## Platinum Sponsors



## Gold Sponsors



# Introduction to R

# What is R?

Open source

Language and environment

Numerical and graphical analysis

Cross platform



# What is R?

Active development  
Large user community  
Modular and extensible  
6700+ extensions



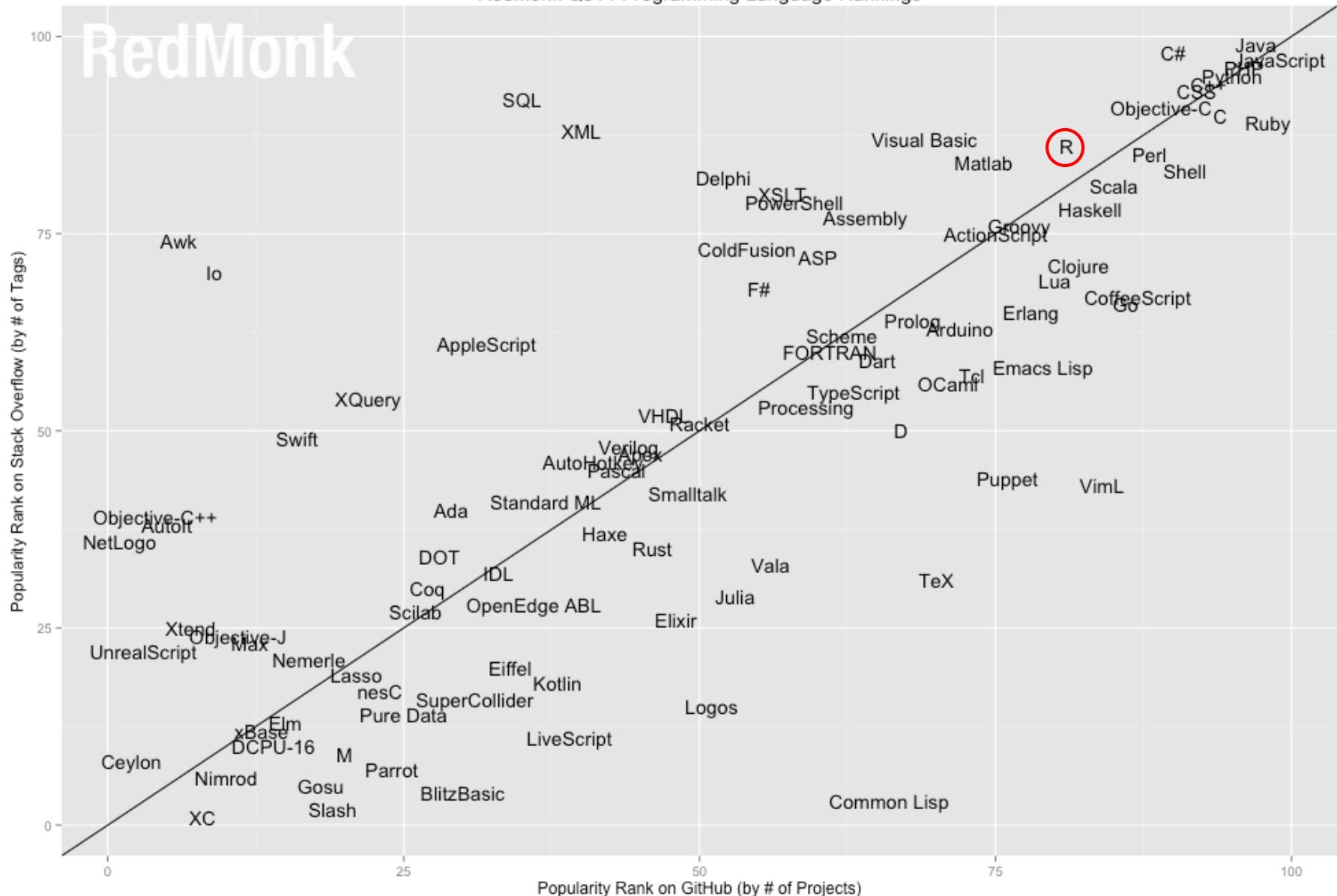
# FREE



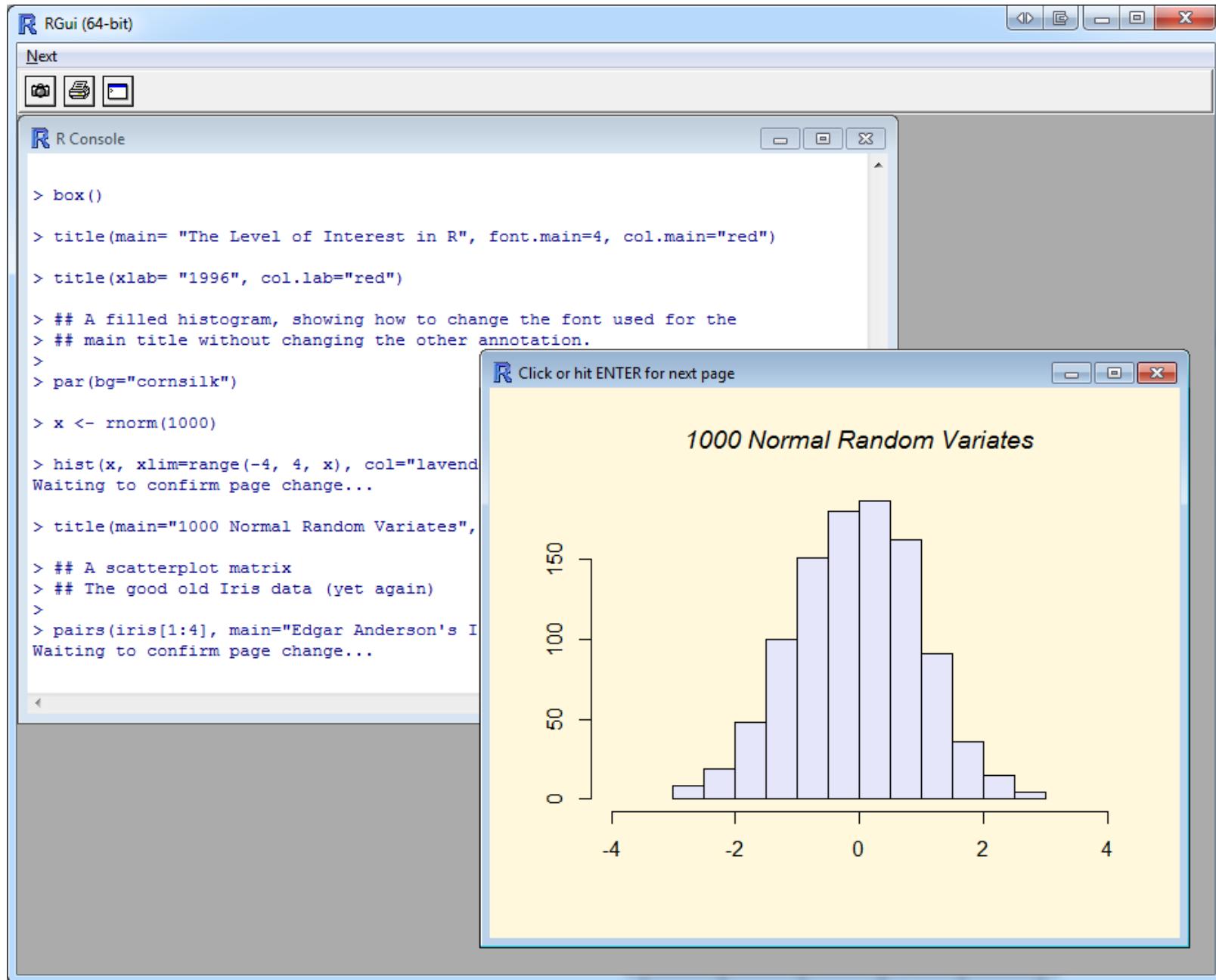
A low-angle photograph of the Statue of Liberty against a clear blue sky. She is shown from the chest up, facing slightly left. Her right arm is raised high, holding a torch aloft. Her left arm is bent, holding a tablet or smartphone that displays the word "FREE".

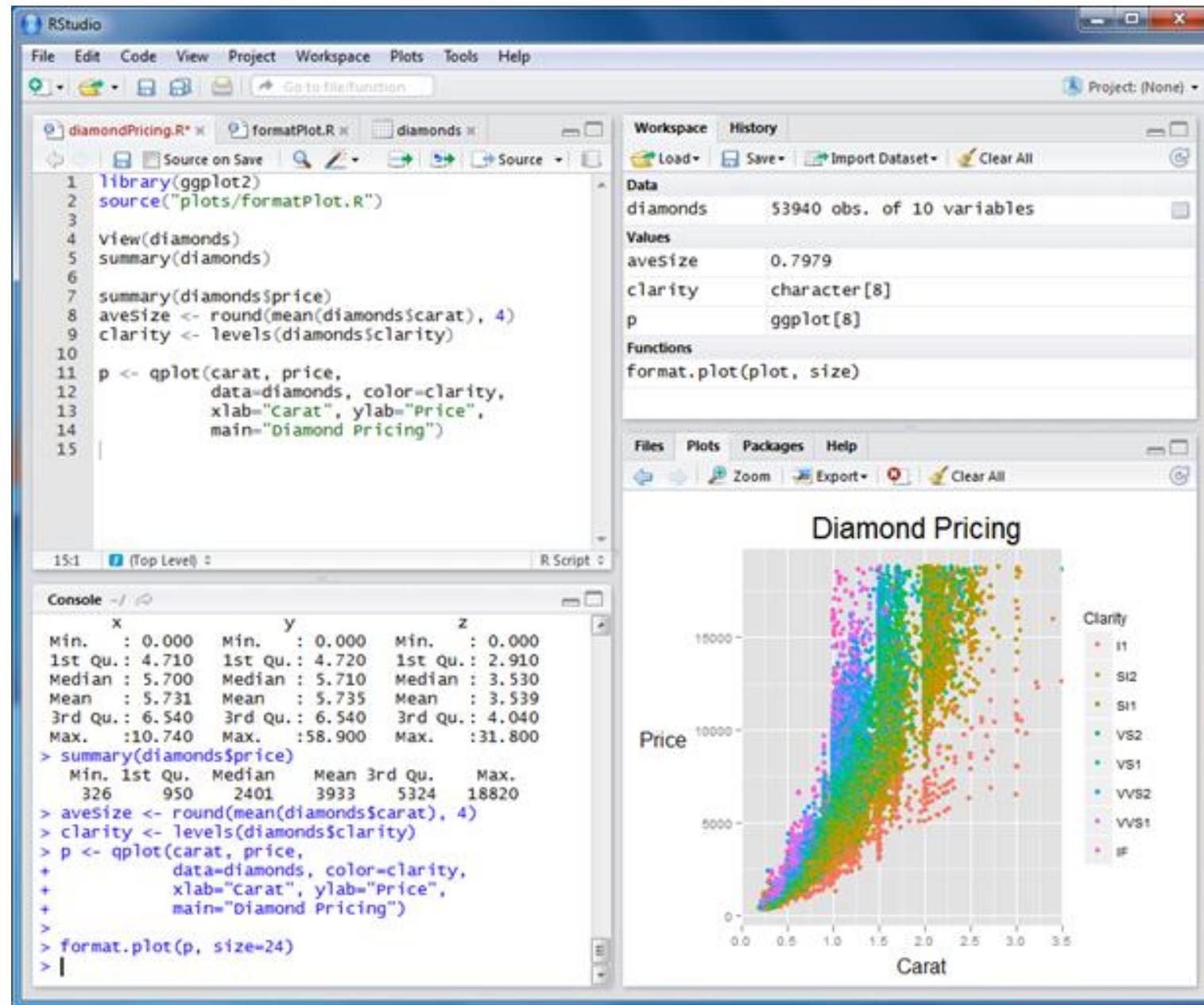
FREE

## RedMonk Q314 Programming Language Rankings



Source: <http://redmonk.com/sogrady/2014/06/13/language-rankings-6-14/>





1-1 - Introduction (Base).R - Microsoft Visual Studio

File Edit View NCrunch Project Debug Team Data Lake Tools Architecture Test ReSharper R Tools Analyze Window Help Matthew Renze

1-1 - Introduction (Base).R Script1.R\*

```
# Create a bar chart
barplot(
  names = df$Name,
  height = df$Value,
  col = "skyblue",
  main = "Hello World",
  xlab = "Name",
  ylab = "Value")

# View the help files
?plot

?barplot
```

R Interactive

```
> # Plot using default parameter order
> plot(df$Name, df$Value)
> # Plot using named parameters
> plot(
+   x = df$Name,
+   y = df$Value)
> # Create a bar chart
> barplot(
+   names = df$Name,
+   height = df$Value,
+   col = "skyblue",
+   main = "Hello World",
+   xlab = "Name",
+   ylab = "Value")
>
```

Variable Explorer

Global Environment

Name	Value	Class	Type
df	3 obs. of 2 variables	data.frame	list
@.Data	List of 2	list	list
@names	chr [1:2] "Name" "Value"	character	character
@row.names	int [1:3] 1 2 3	integer	integer
@.S3Class	"data.frame"	character	character
Name	Factor w/ 3 levels "a","b","c": 1 2 3	factor	integer
Value	num [1:3] 1 2 3	numeric	double

Variable Explorer R History

R Plot

Hello World

Value

a b c

Name

Solution Explorer R Plot R Help

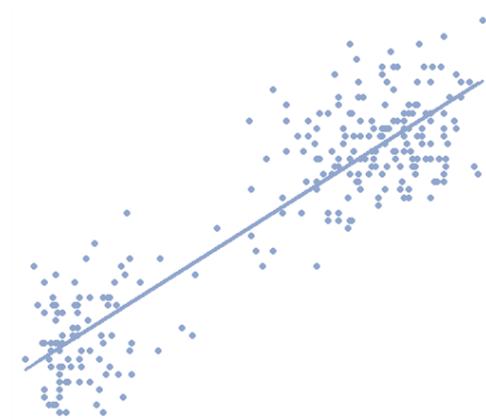
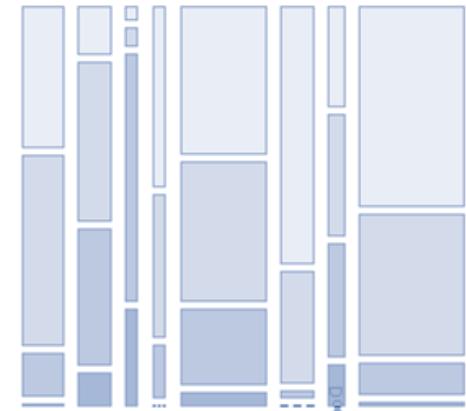
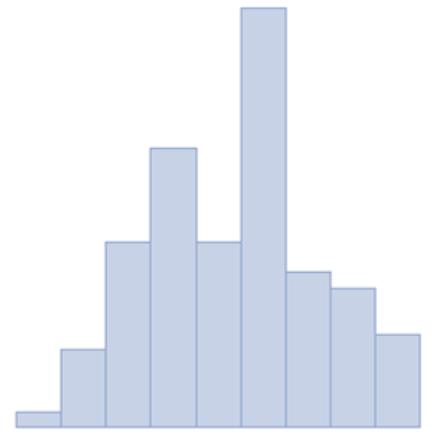
Ready Ln 29 Col 1 Ch 1 INS Cloud develop ✓

# Code Demo

# Intro to Data Visualization

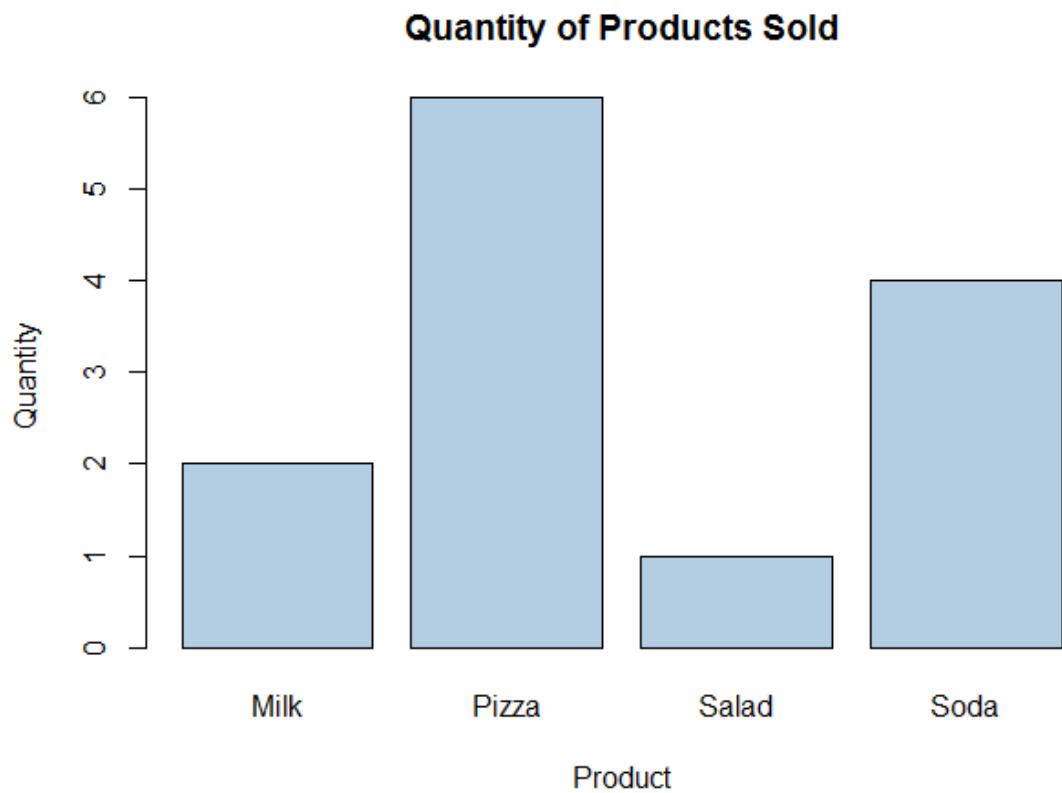
# Data Visualization

Visual data representation  
For human pattern recognition  
Map dimensions to visual



# Data Visualization

ID	Date	Customer	Product	Quantity
1	2015-08-27	John	Pizza	2
2	2015-08-27	John	Soda	2
3	2015-08-27	Jill	Salad	1
4	2015-08-27	Jill	Milk	1
5	2015-08-28	Miko	Pizza	3
6	2015-08-28	Miko	Soda	2
7	2015-08-28	Sam	Pizza	1
8	2015-08-28	Sam	Milk	1



# Statistical Terms

Observations

Variables

Qualitative variable

Quantitative variable

ID	Date	Customer	Product	Quantity
1	2015-08-27	John	Pizza	2
2	2015-08-27	John	Soda	2
3	2015-08-27	Jill	Salad	1
4	2015-08-27	Jill	Milk	1
5	2015-08-28	Miko	Pizza	3
6	2015-08-28	Miko	Soda	2
7	2015-08-28	Sam	Pizza	1
8	2015-08-28	Sam	Milk	1

# Types of Data Visualizations

Number of variables  
Type of variable(s)

Number of Variables

One  
Categorical  
Variable

One  
Numeric  
Variable

Two  
Categorical  
Variables

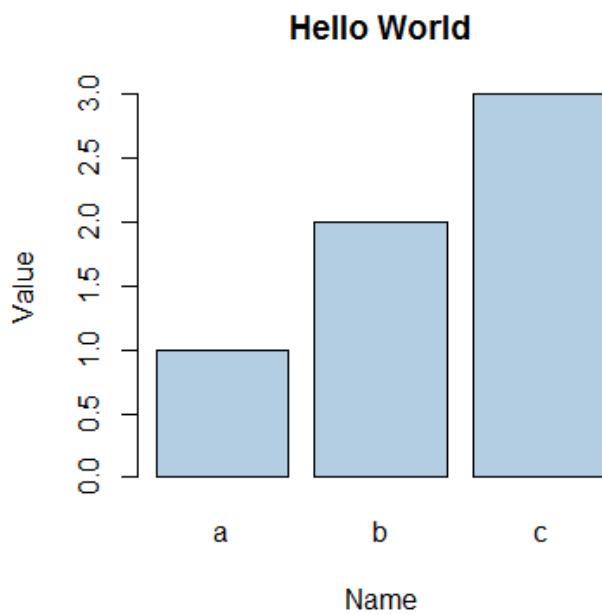
Categorical  
& Numeric  
Variable

Two  
Numeric  
Variables

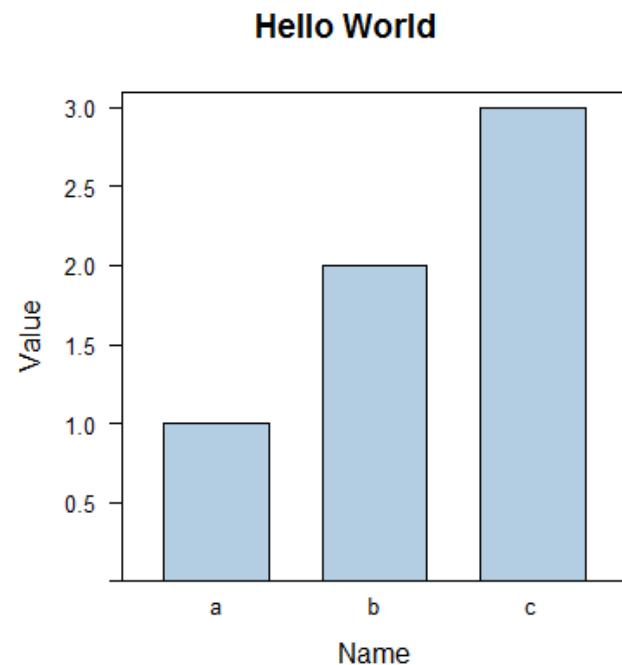
Many  
Variables

Type of Variable(s)

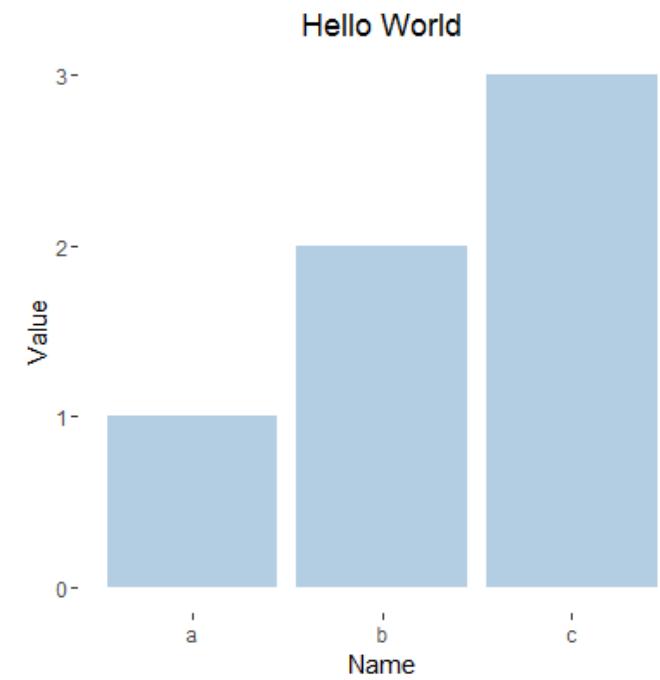
# Plotting Systems in R



Base



Lattice

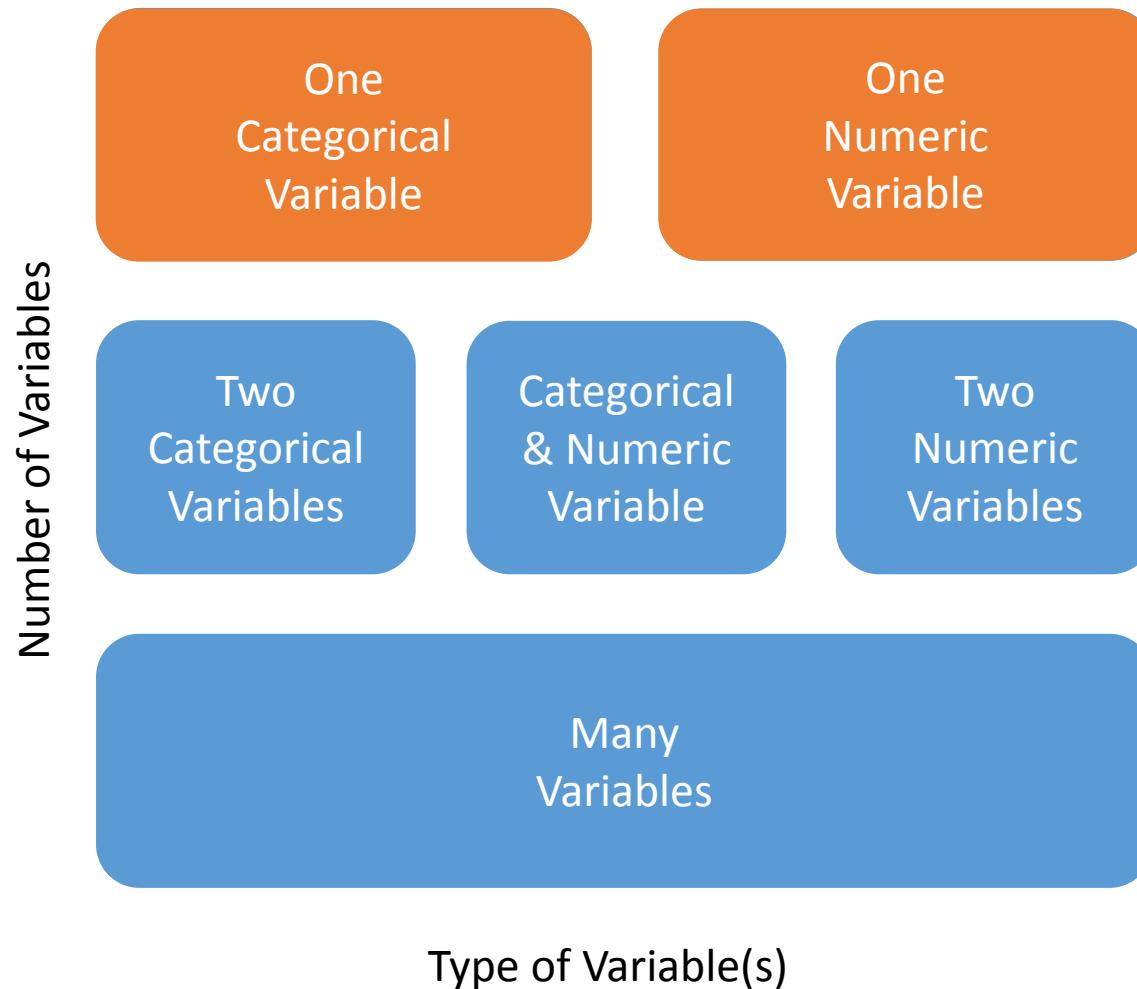


ggplot2

# Code Demo

# Visualizing One Variable

# Visualizing One Variable



# Visualizing One Categorical Variable

Frequency  
Proportion

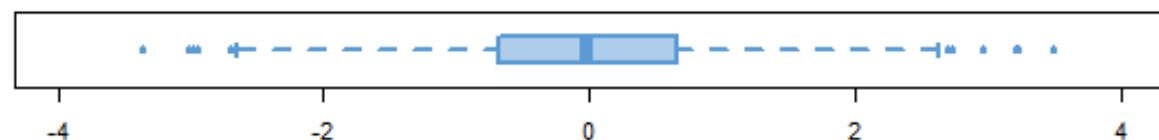
Movies by Genre		
Genre	Frequency	Percentage
Action	612	9%
Adventure	496	7%
Animation	168	2%
Comedy	1281	18%
Drama	1570	22%
Horror	269	4%
...	...	...

# Visualizing One Numeric Variable

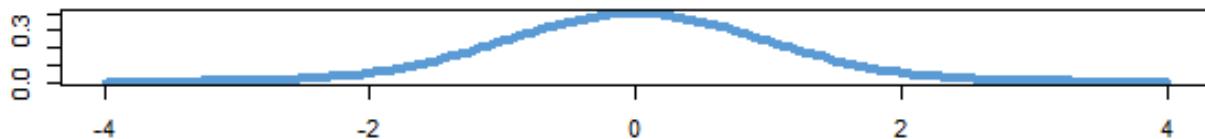
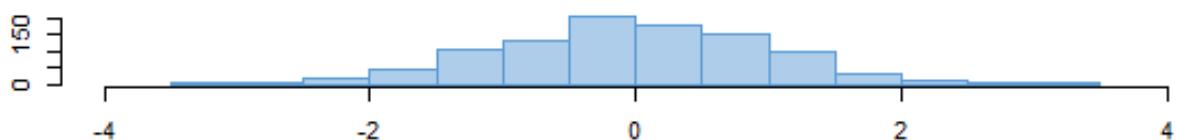
Location



Spread



Shape



# Open Movies Database

Movies								
Title	Year	Rating	Runtime (minutes)	Genre	Critic Score	Box Office	Awards	Distribution
The Whole Nine Yards	2000	R	98	Comedy	45%	\$57.3M	No	Domestic
Cirque du Soleil	2000	G	39	Family	45%	\$13.4M	Yes	Domestic
Gladiator	2000	R	155	Action	76%	\$187.3M	Yes	International
Dinosaur	2000	PG	82	Family	65%	\$135.6M	Yes	Domestic
Big Momma's House	2000	PG-13	99	Comedy	30%	\$0.5M	Yes	International



PROD. NO.  
SCENE

TAKE

ROLL





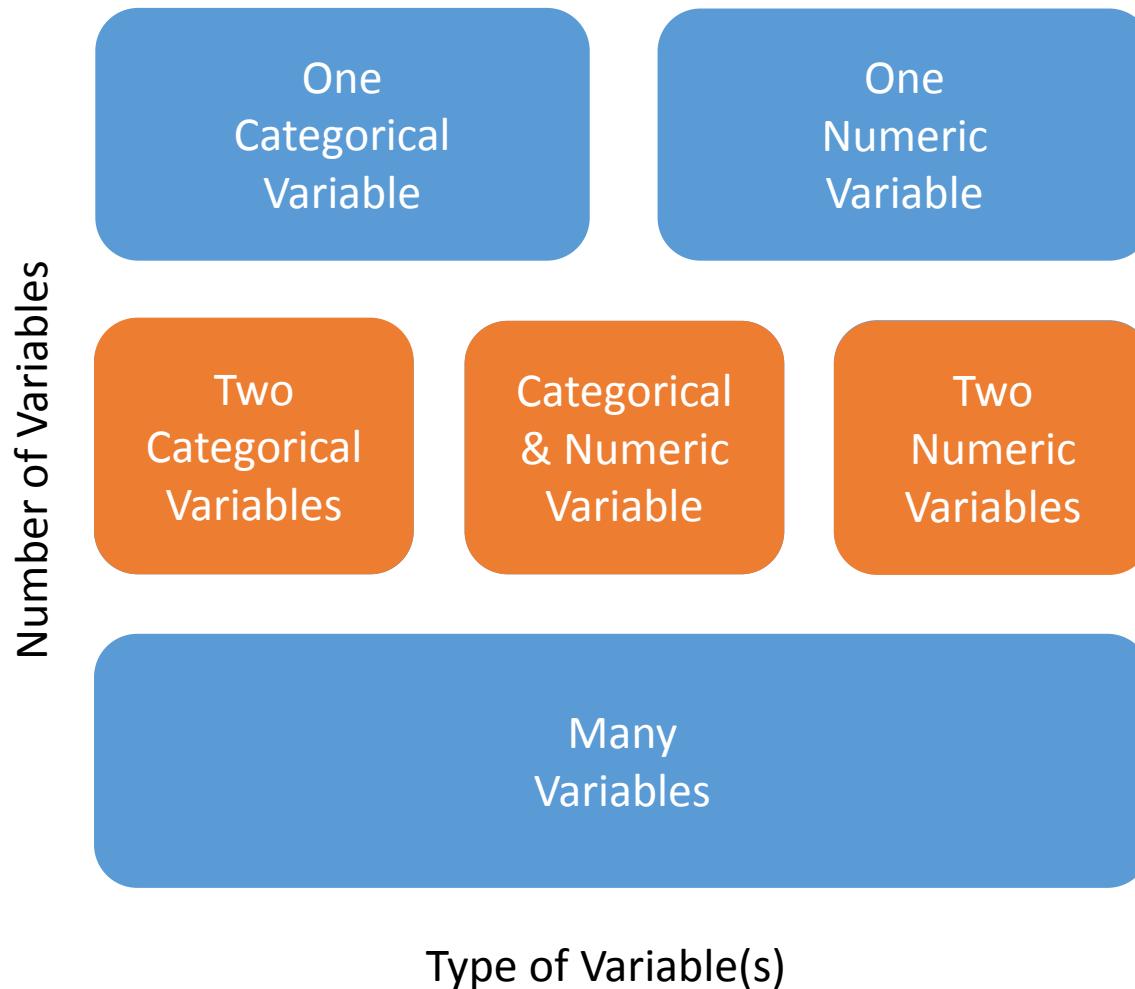
1. How does the count of movies compare by rating category?
2. What proportion of movies have won awards?
3. What does the distribution of movie runtimes look like?

# Code Demo



# Visualizing Two Variables

# Visualizing Two Variables



# Visualizing Two Categorical Variables

Joint frequency

Marginal frequency

Relative frequency

Movies by Genre and Rating					
Genre	G	PG	PG-13	R	Total
Action	2	70	311	229	612
Adventure	44	179	209	64	496
Animation	43	111	8	6	168
Comedy	45	258	472	506	1218
Drama	12	136	586	836	1570
Family	38	181	10	1	230
...	...	...	...	...	...
Total	230	1207	2686	3058	7181

# Visualizing Two Categorical Variables

Joint frequency

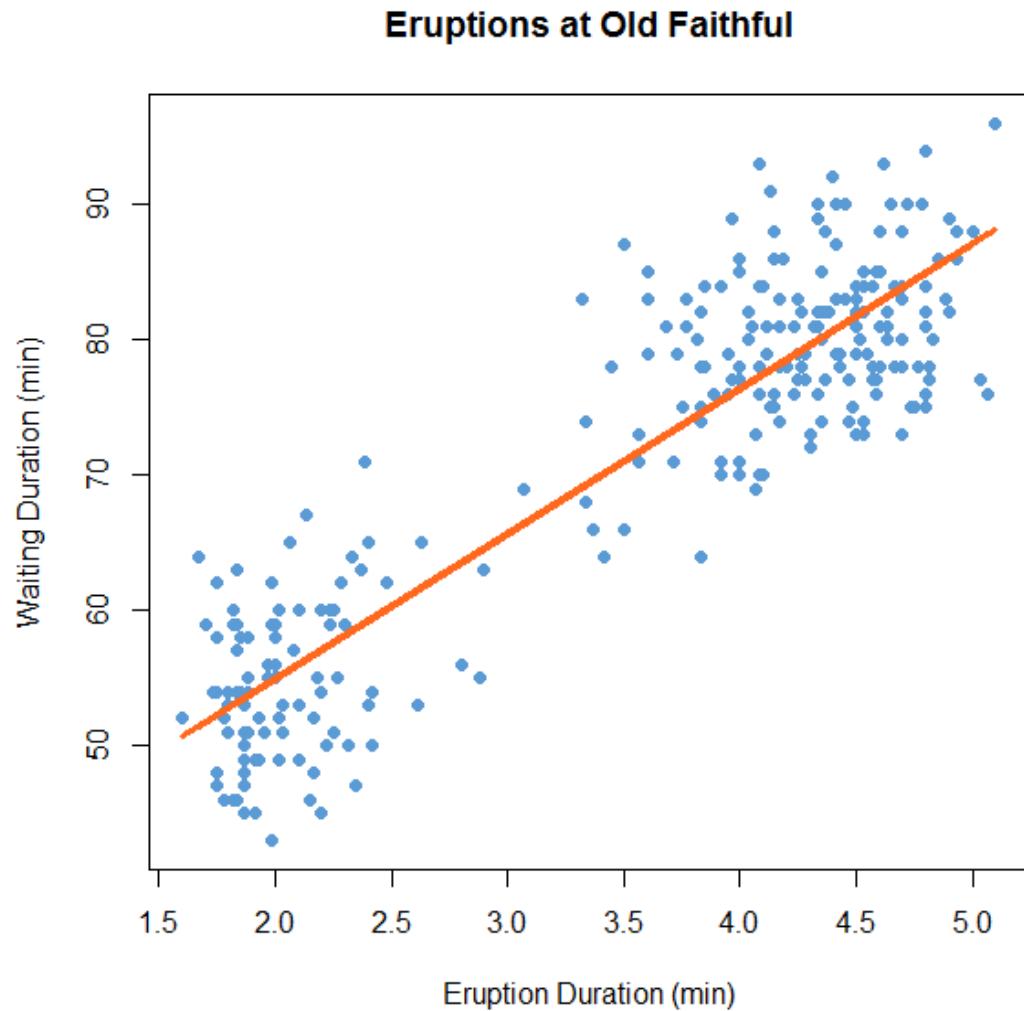
Marginal frequency

Relative frequency

Movies by Genre and Rating					
Genre	G	PG	PG-13	R	Total
Action	0.001	0.010	0.043	0.032	0.086
Adventure	0.006	0.025	0.029	0.009	0.069
Animation	0.006	0.015	0.001	0.001	0.023
Comedy	0.006	0.036	0.066	0.070	0.170
Drama	0.002	0.019	0.082	0.116	0.219
Family	0.005	0.025	0.001	0.001	0.033
...	...	...	...	...	...
Total	0.032	0.168	0.374	0.426	1.000

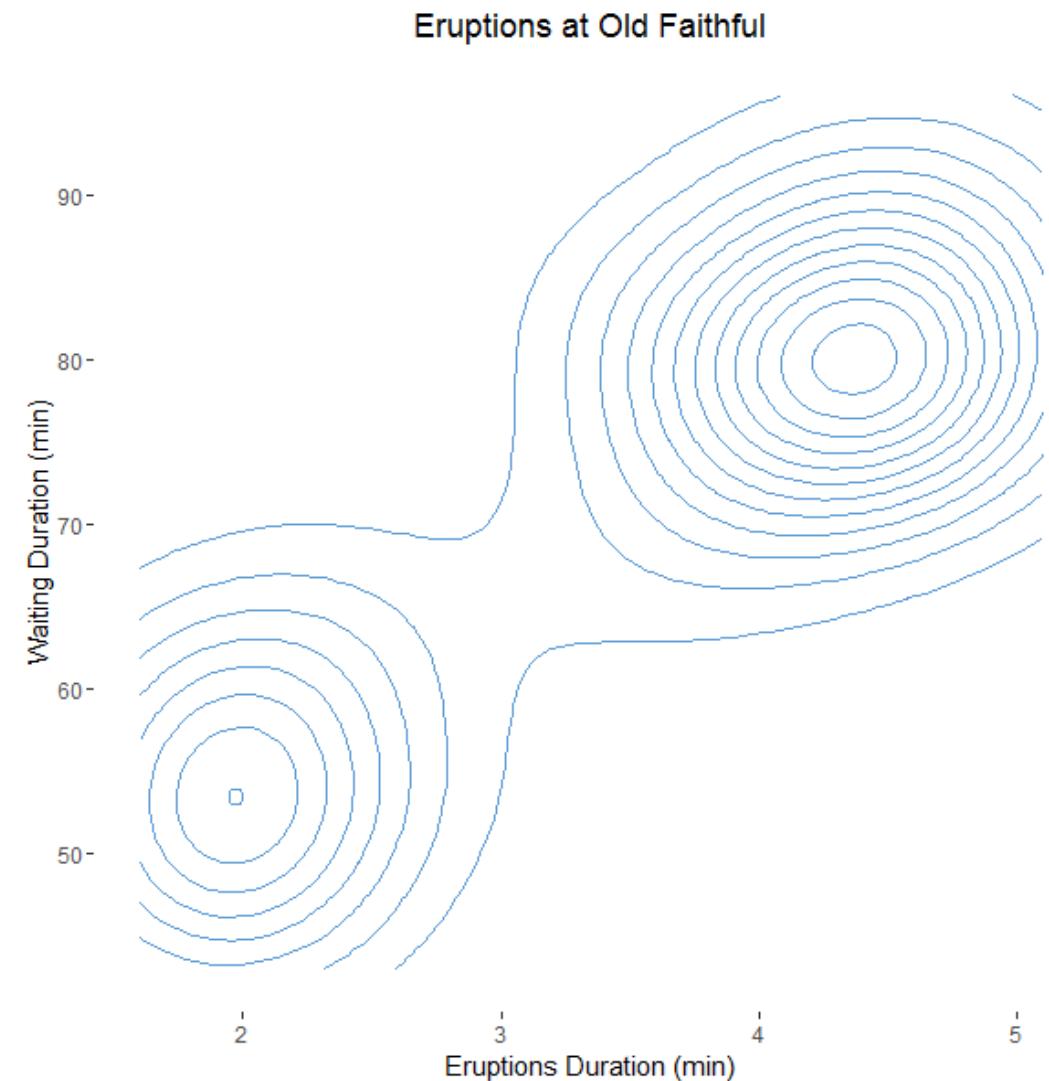
# Visualizing Two Numeric Variables

Relationship  
Correlation  
Distribution



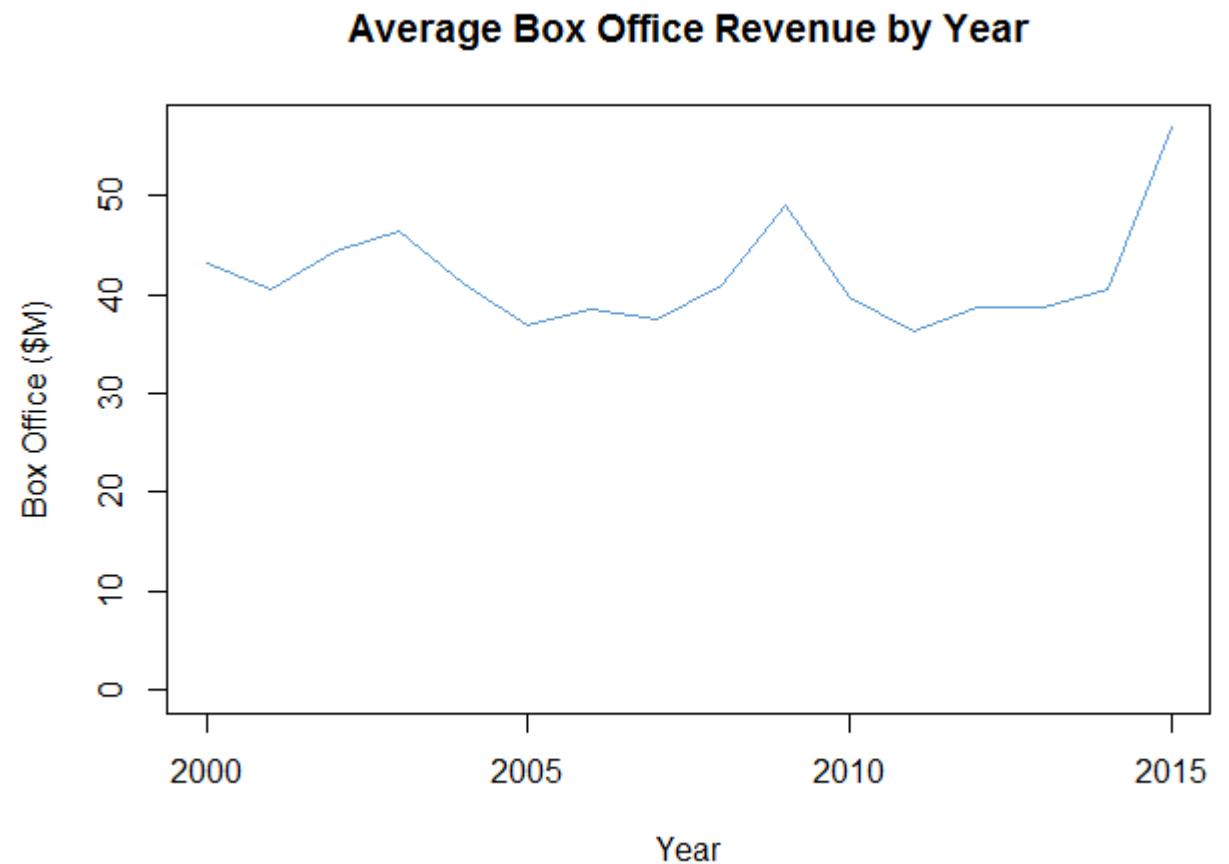
# Visualizing Two Numeric Variables

Relationship  
Correlation  
Distribution



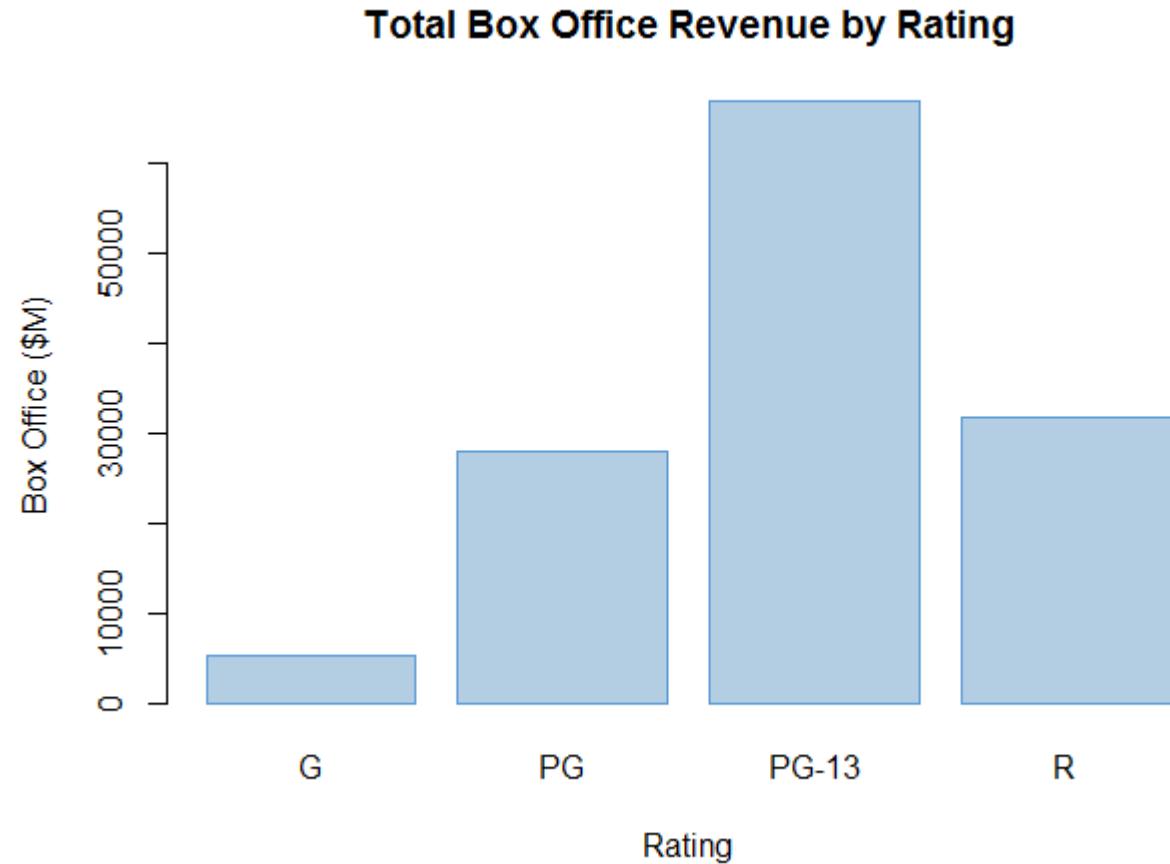
# Visualizing Time Series Data

Values over time  
Rate of change



# Visualizing a Numeric Variable Grouped by a Categorical Variable

Aggregate  
Grouped  
Comparison





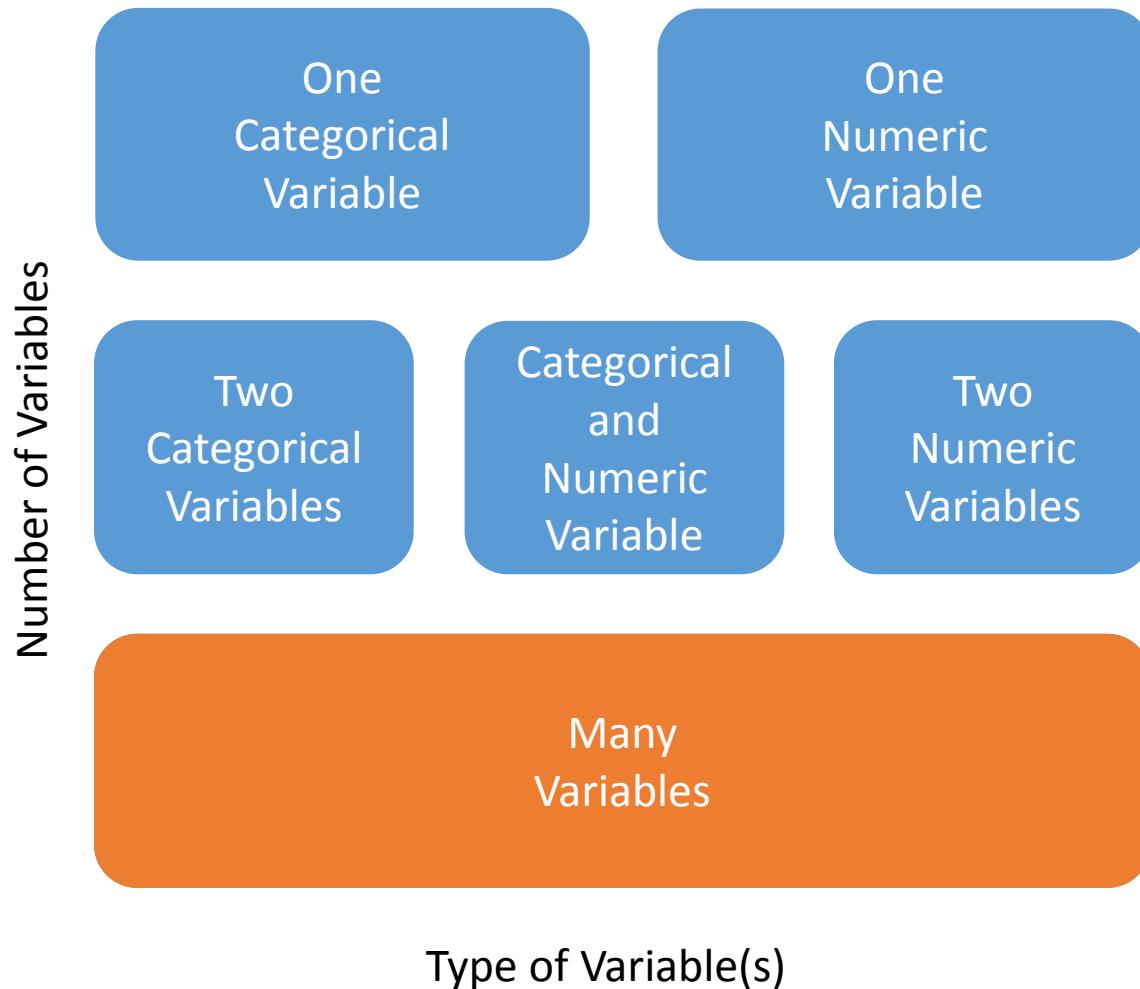
1. How many award-winning movies by rating category?
2. How are runtime and box-office revenue related?
3. Has the average box-office revenue changed over time?
4. How does box office compare by rating category

# Code Demo

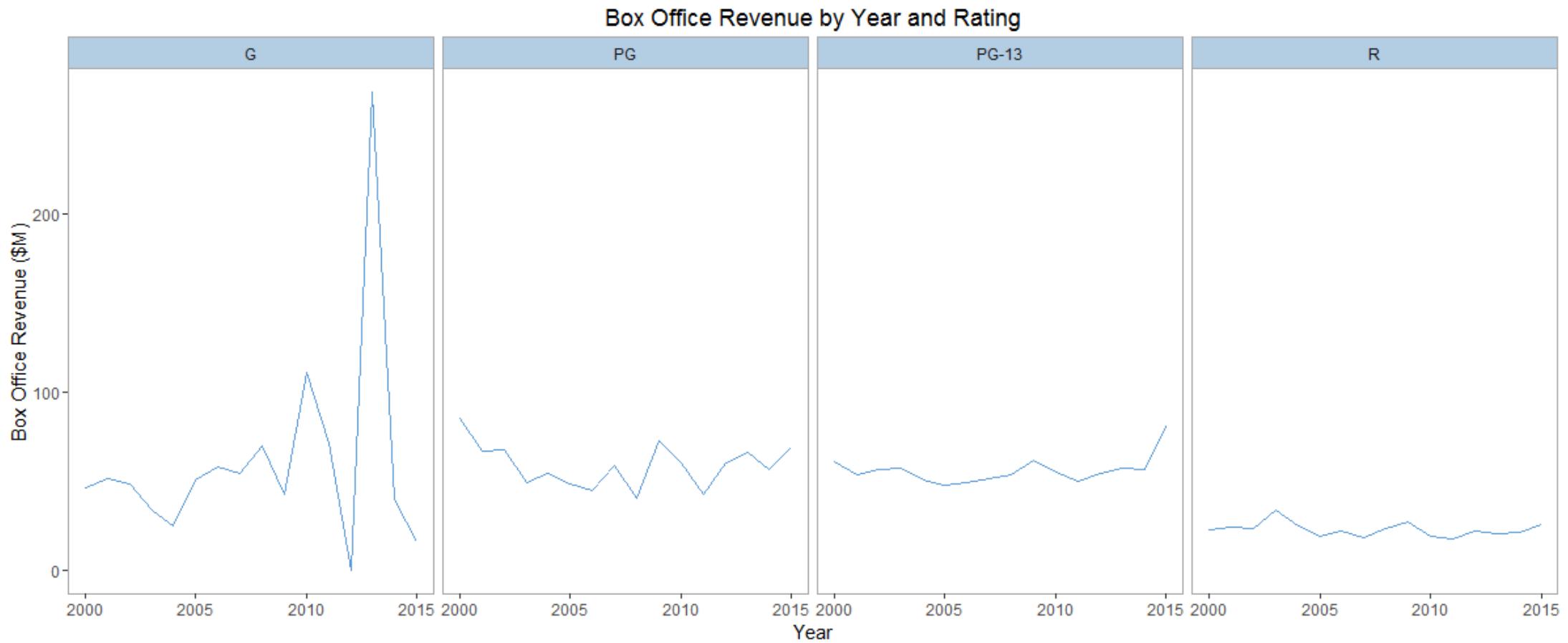


# Visualizing Many Variables

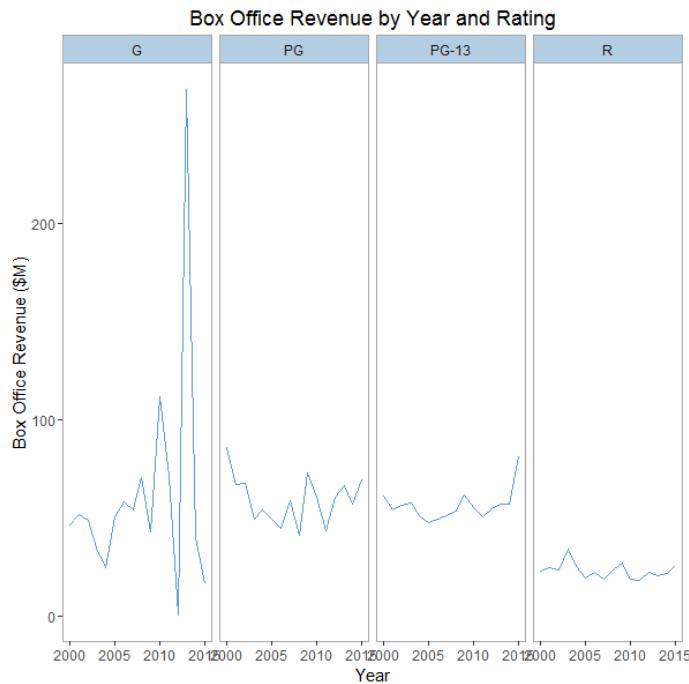
# Visualizing Many Variables



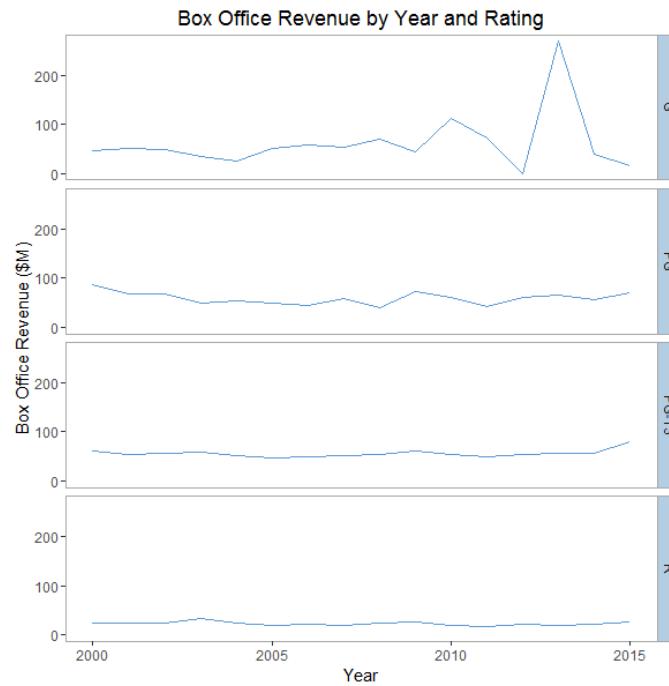
# Faceting



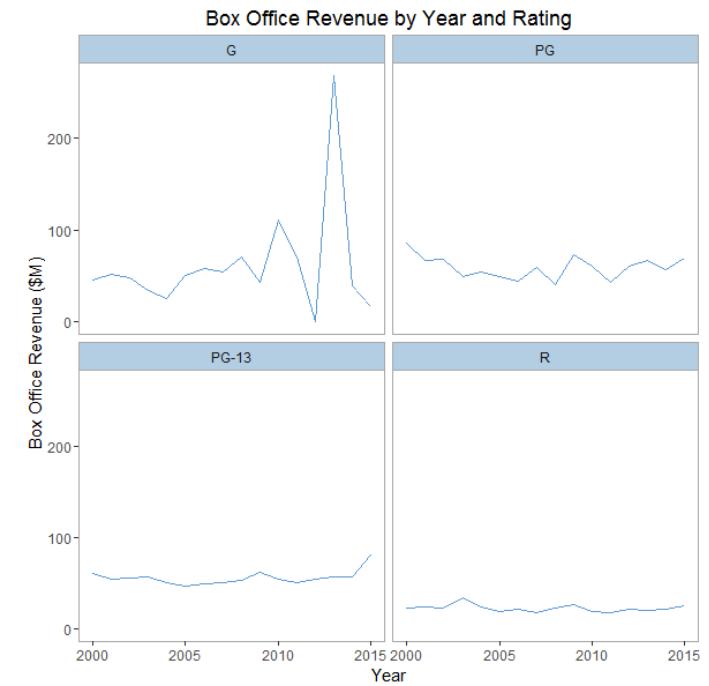
# Faceting Orientation



Horizontal

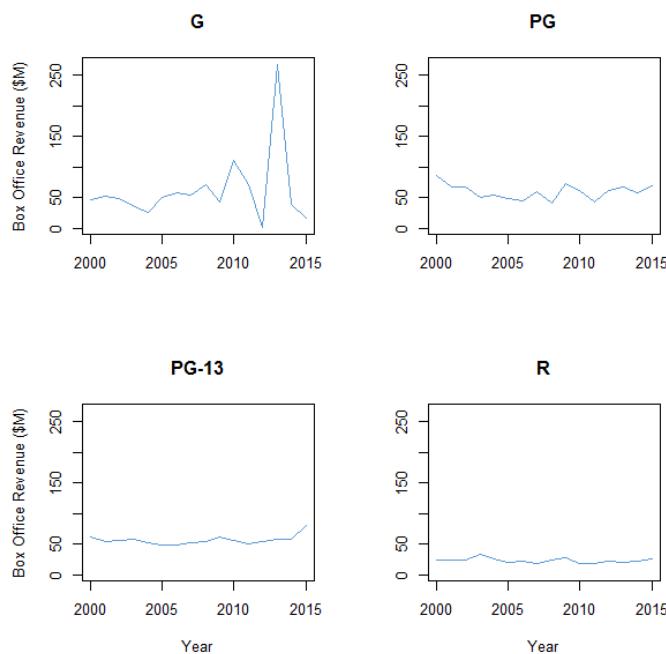


Vertical

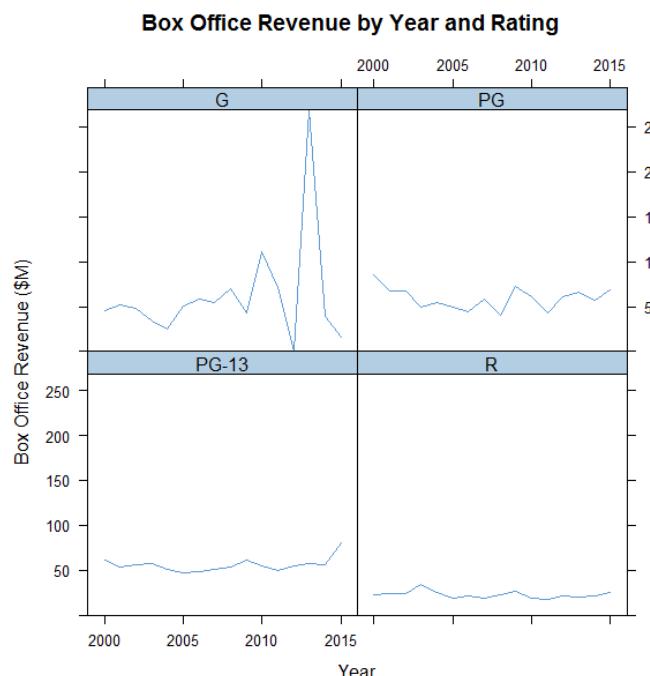


Wrapped

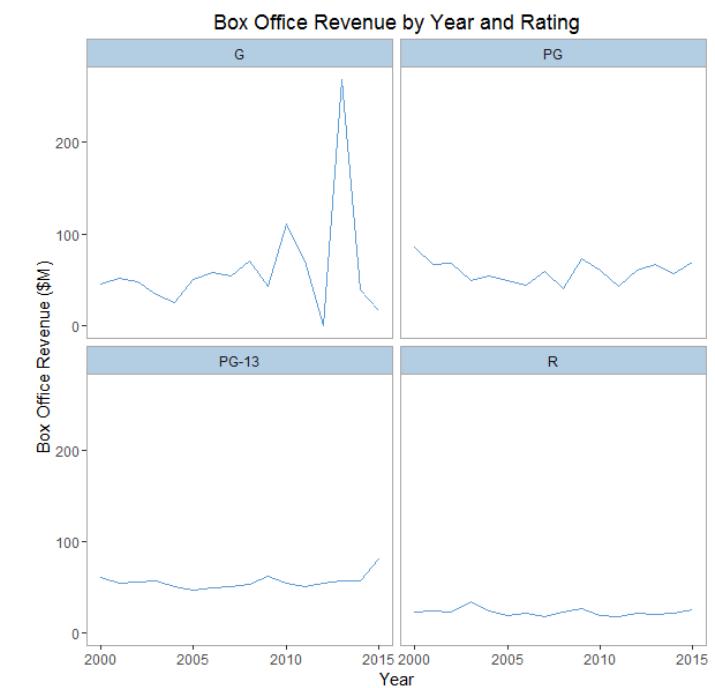
# Faceting



Base



Lattice



ggplot2

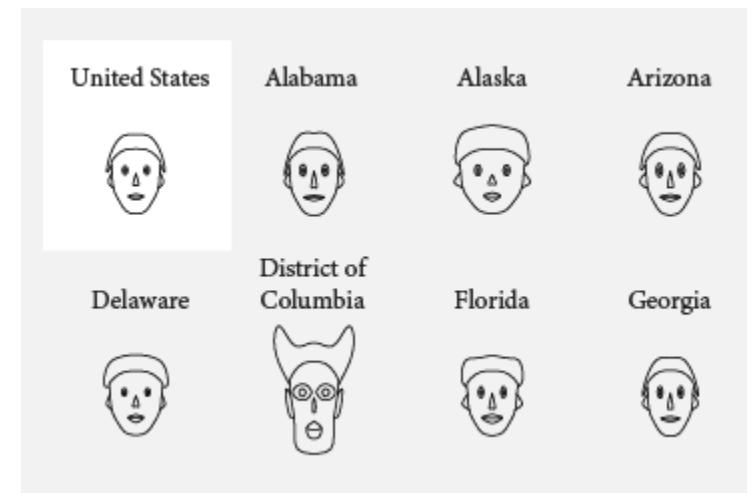
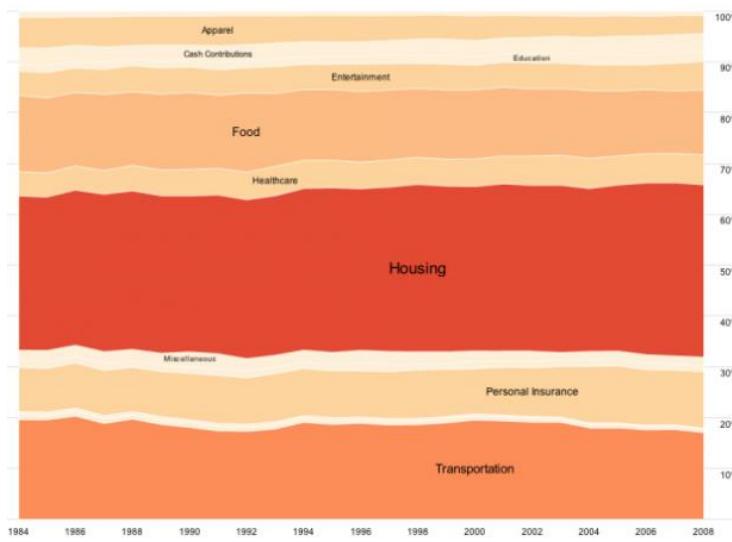
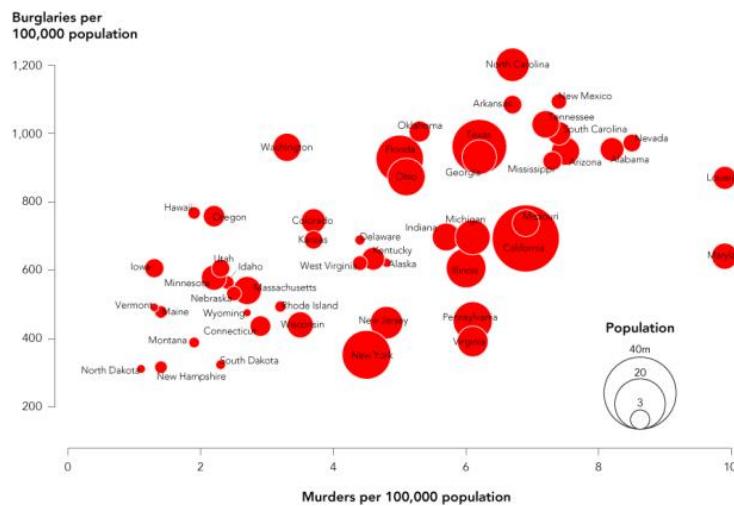
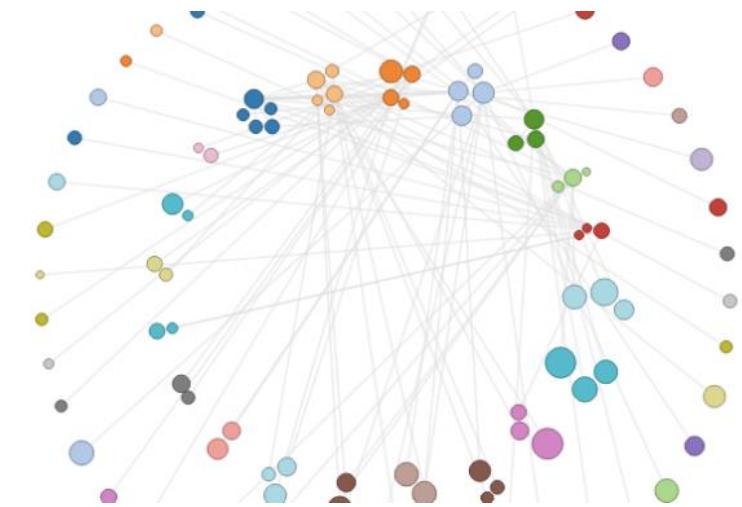
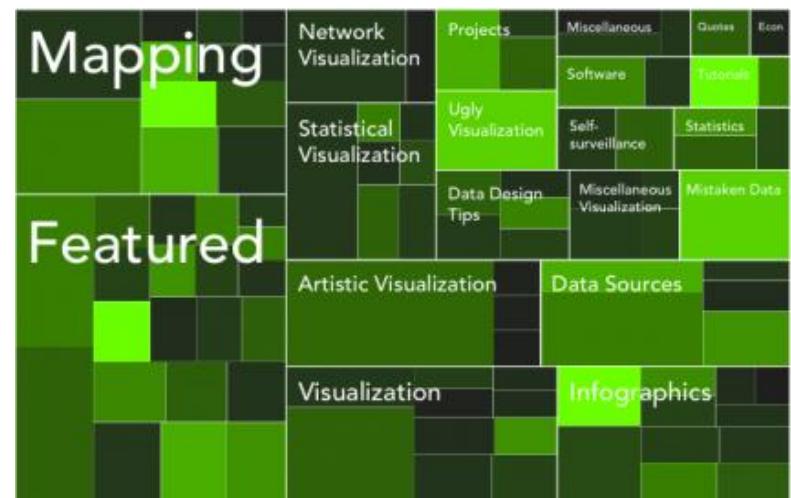
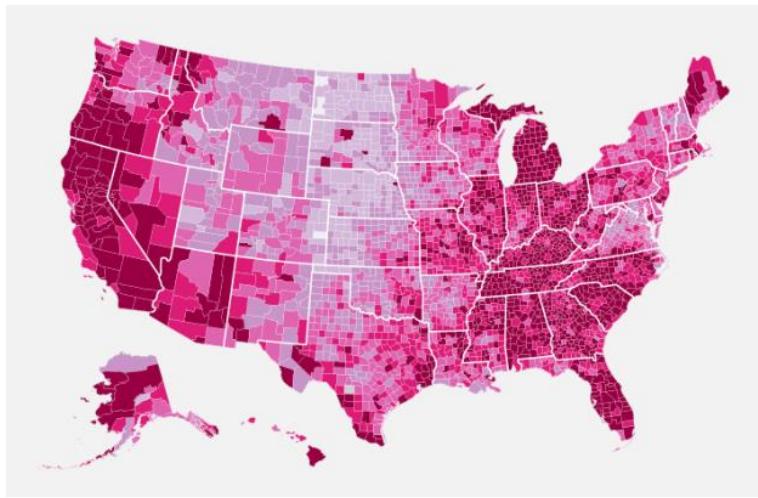


1. How many movies based on rating, award, and dist.?
2. How does box office compare across awards and rating?
3. How does critic score relate to box office by rating category?
4. How has box office revenue changed over time by rating?
5. How do runtime and critic score relate to box office?

# Code Demo



# Visualizing Other Types of Data



Source: Nathan Yau ([www.flowingdata.com](http://www.flowingdata.com))



1. What is the box office revenue for films by country?
2. What is the box office revenue by continent and country?
3. Which actors have acted most frequently with one another?
4. Which movie plot words are associated with revenue?

# Code Demo

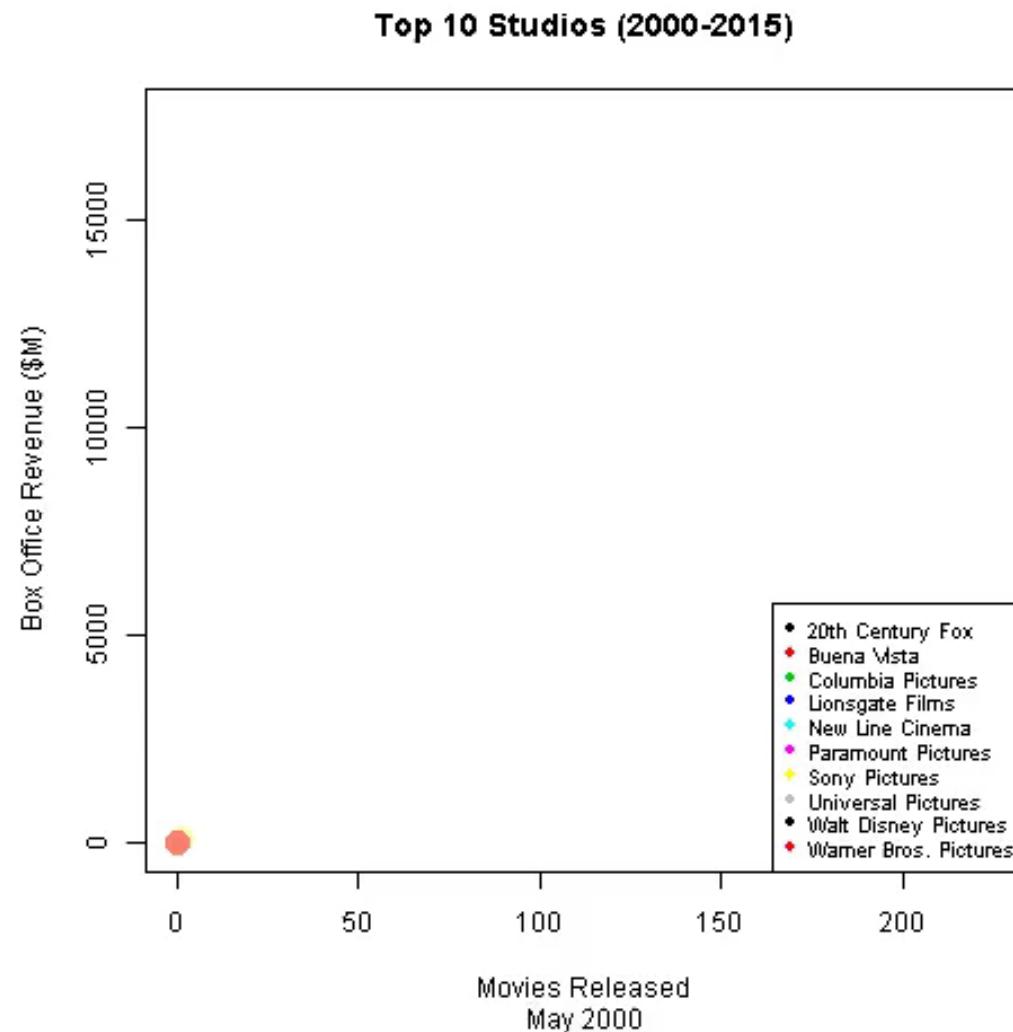


# Beyond the Basics

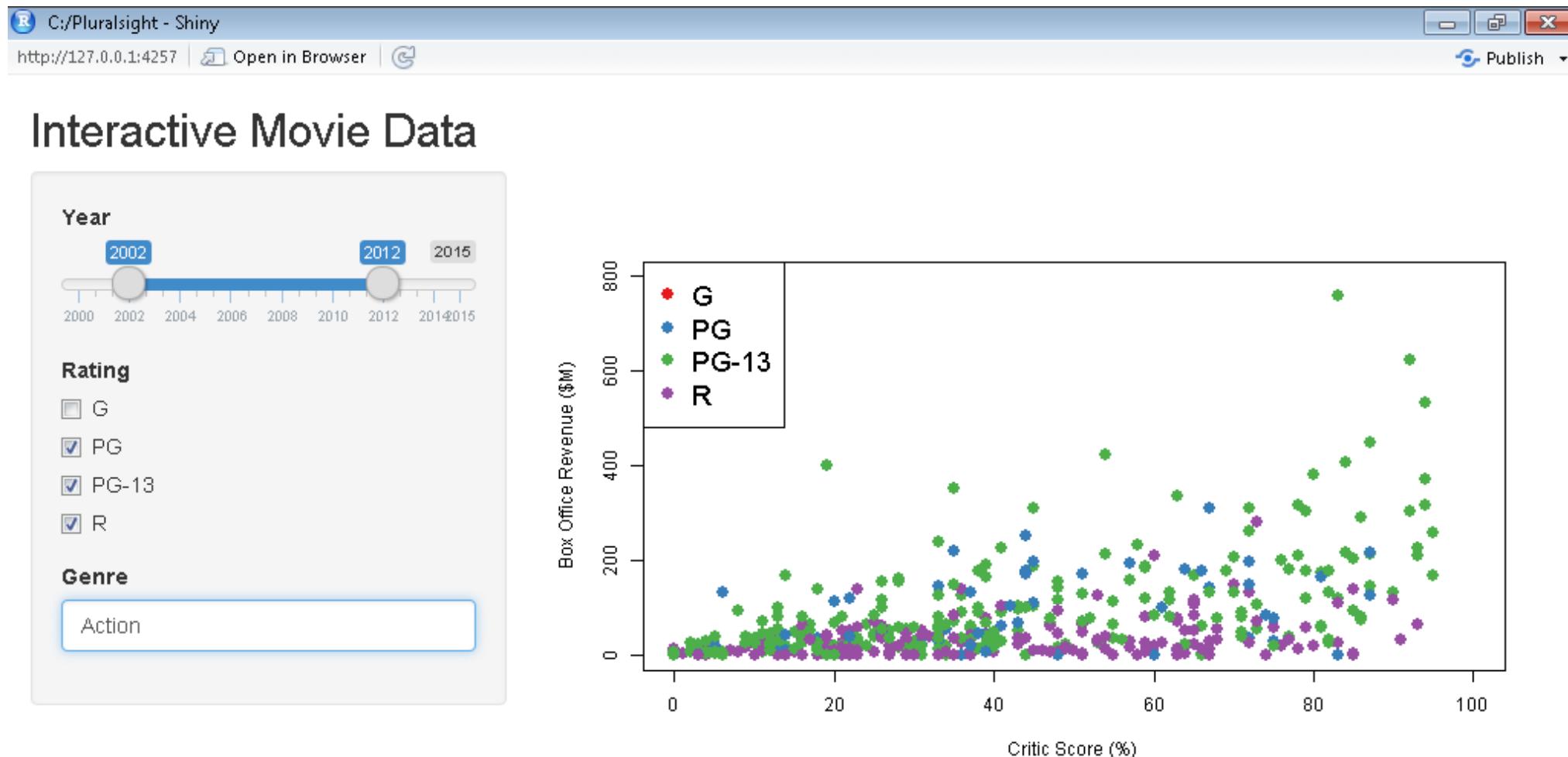


This is just the tip of the iceberg!

# Animating Data Visualizations



# Building Interactive Data Visualizations





1. How have the top 10 studios grown over the past 15 years?
2. Allow me to explore the data to ask any question I might dream up on a random Tues.

... and I want it done today.

# Code Demo



# Where to Go Next...

R website: <http://www.cran.r-project.org>

R Studio: <http://www.rstudio.com>

Revolutions: <http://blog.revolutionanalytics.com>

Flowing Data: <http://flowingdata.com>

R-Blogger: <http://www.r-bloggers.com>

R-Seek: <http://rseek.org>

# My Website

Articles

Presentations

Source Code

Videos

Workshops

Matthew Renze

Home Articles Courses Presentations Software About Contact

## News

### 2016-06-17 - Tinnitus Conference Presentation

I spoke at the 24th International Tinnitus Conference at the University of Iowa Medical College yesterday. The feedback I received on my presentation was very positive. Thank you to Dr. Rich Tyler for inviting me to speak at my first medical conference.



### 2016-06-11 - Invitation to Speak at Dev Up Conference

I received an invitation to speak at the new Dev Up Conference (formerly St. Louis Days of .NET). This year, I'll be presenting my Clean Architecture talk and my new HoloLens talk.



### 2016-06-02 - Invitation to Speak at AIM | hdc

I've been invited to speak at AIM | hdc again this year. This time, I will be presenting a full-day version of my Practical Data Science with R workshop. I'm looking forward to seeing the whole HDC crew again this year.



### 2016-05-20 - Nebraska Code Presentations

Both my workshop on Practical Data Science with R and my presentation on Clean Architecture at Nebraska Code went extremely well. I also had a great time seeing with all of my friends at the conference.



[www.matthewrenze.com](http://www.matthewrenze.com)



PLURALSIGHT

Exploratory Data Analysis with R  
Beginning Data Visualization with R  
Multivariate Data Visualization with R  
Mastering Data Visualization with R

Mastering Data Visualization with R



**Matthew Renze**

SOFTWARE CONSULTANT

@matthewrenze www.matthewrenze.com



[www.pluralsight.com/authors/matthew-renze](http://www.pluralsight.com/authors/matthew-renze)

# Conclusion

# Conclusion

Introduction to R

Intro to Data Visualization

Types of Data Visualizations

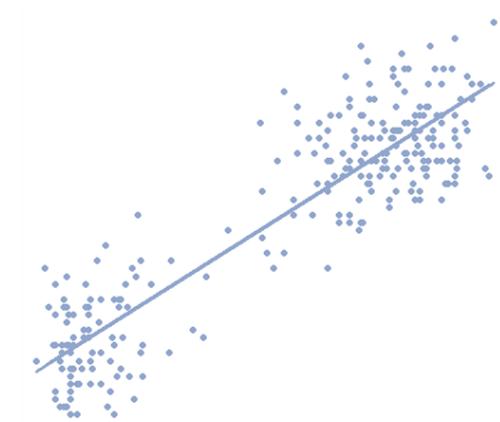
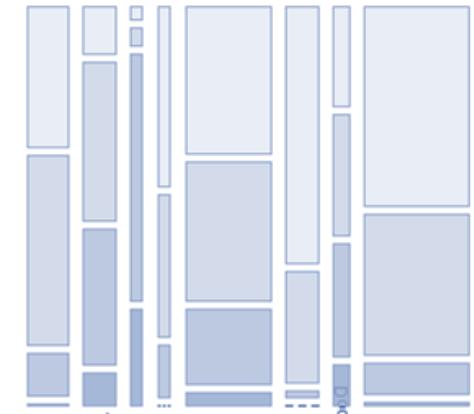
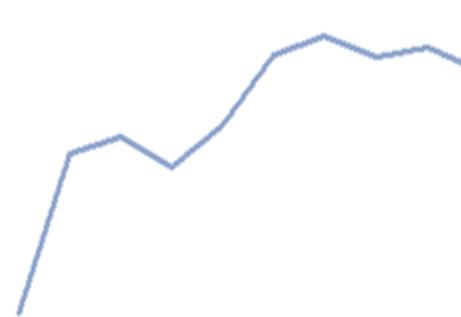
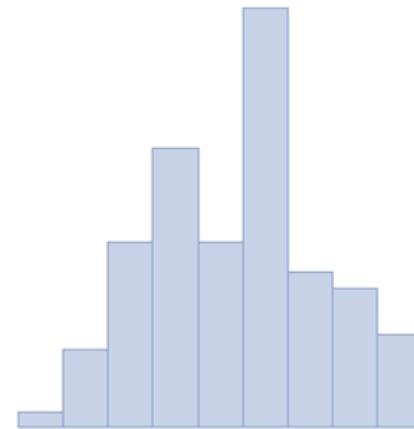
- Visualizing One Variable

- Visualizing Two Variables

- Visualizing Many Variables

- Visualizing Other Types of Data

Beyond the Basics

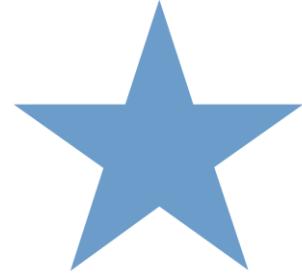


# Feedback

Feedback is very important to me!

One thing you liked?

One thing I could improve?



# Contact Info

Matthew Renze

Data Science Consultant

Renze Consulting

Twitter: [@matthewrenze](https://twitter.com/matthewrenze)

Email: [matthew@matthewrenze.com](mailto:matthew@matthewrenze.com)

Website: [www.matthewrenze.com](http://www.matthewrenze.com)



Thank You! : )