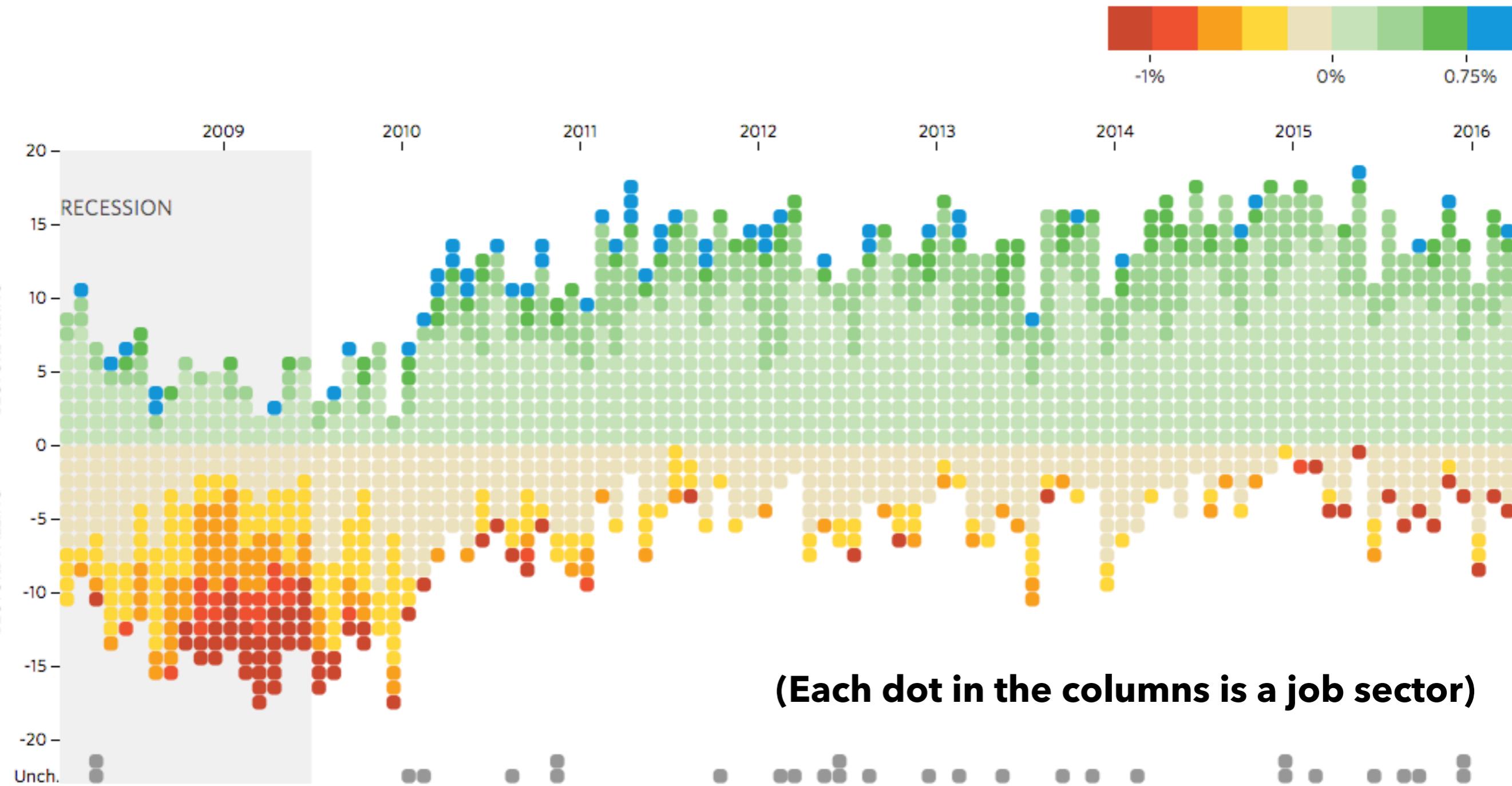


What are the marks? What are the visual channels?

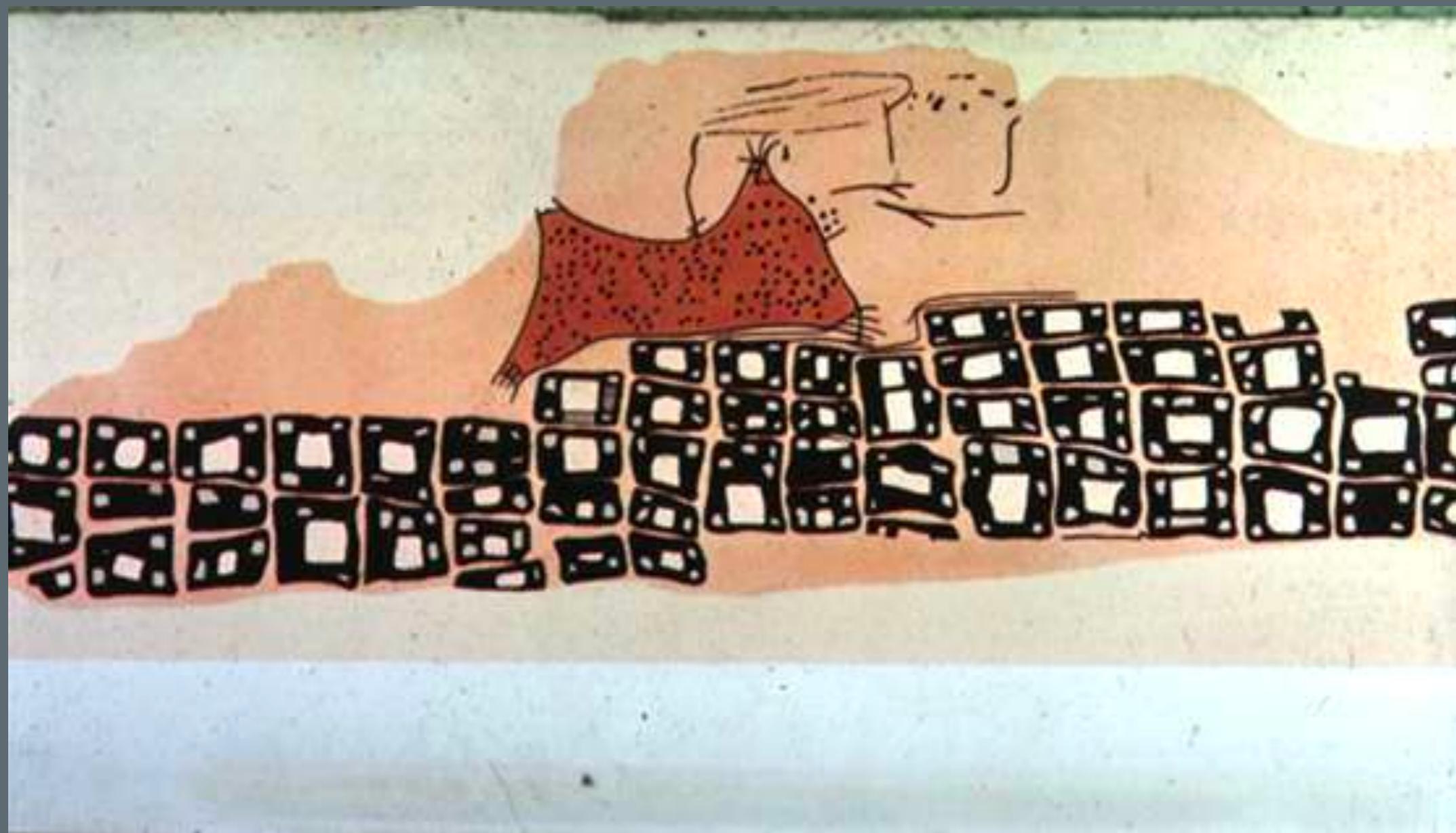


Brief History of Data Visualization



6000? BCE

Çatalhöyük Map?

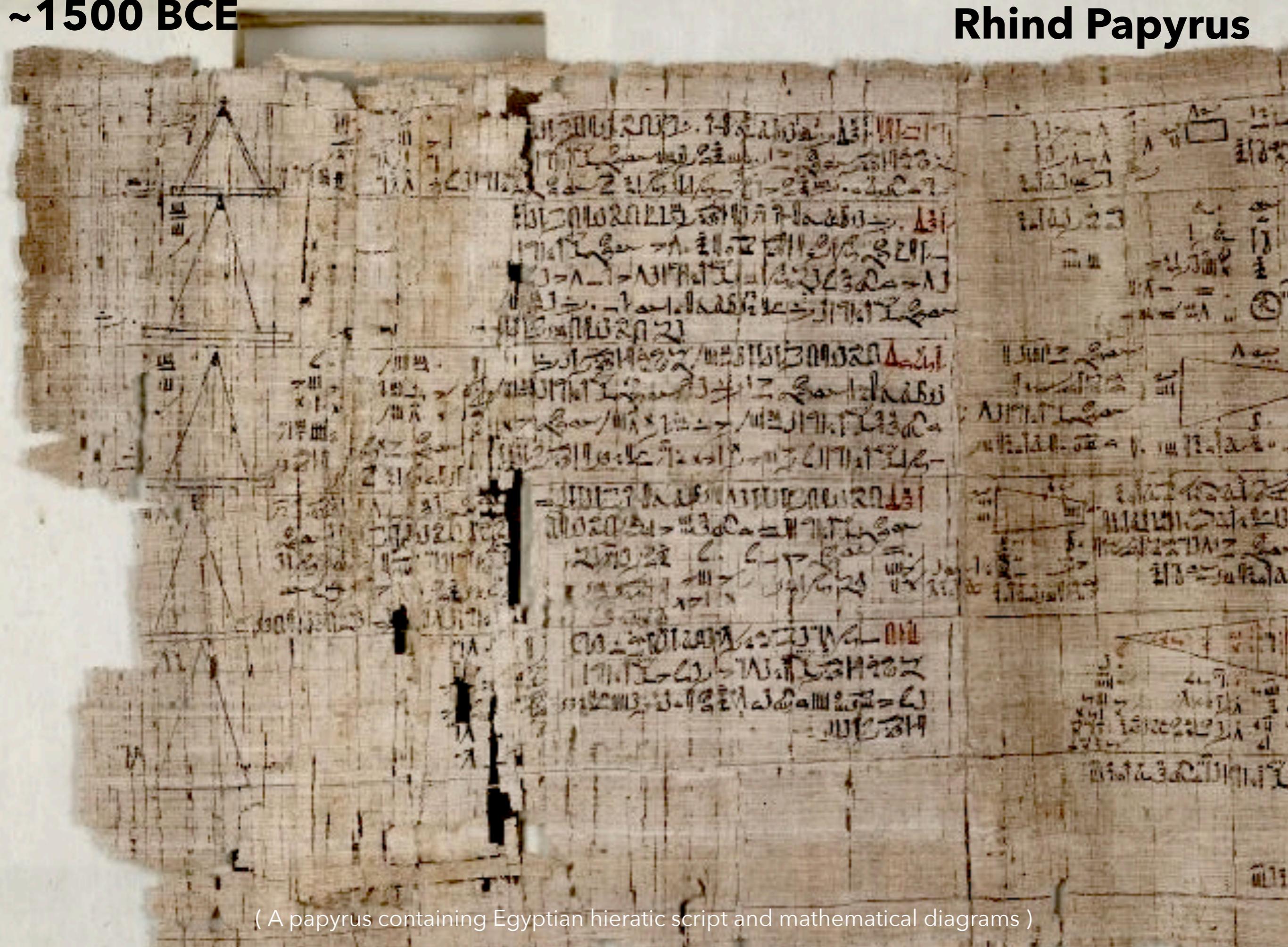


(A "map" depicting a village in brown pigments; may actually be a picture of a leopard skin)

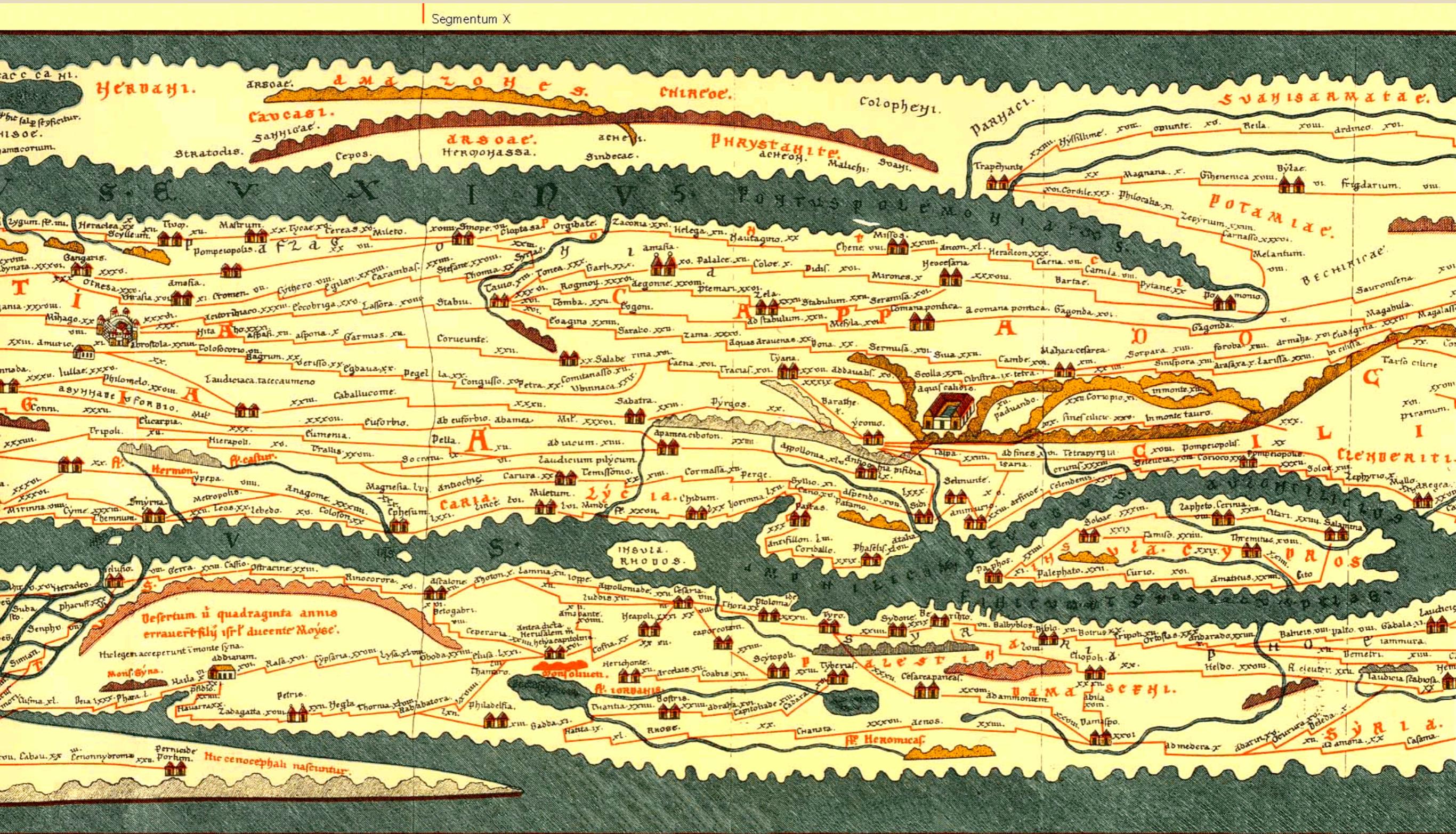


~1500 BCE

Rhind Papyrus



(A papyrus containing Egyptian hieratic script and mathematical diagrams)

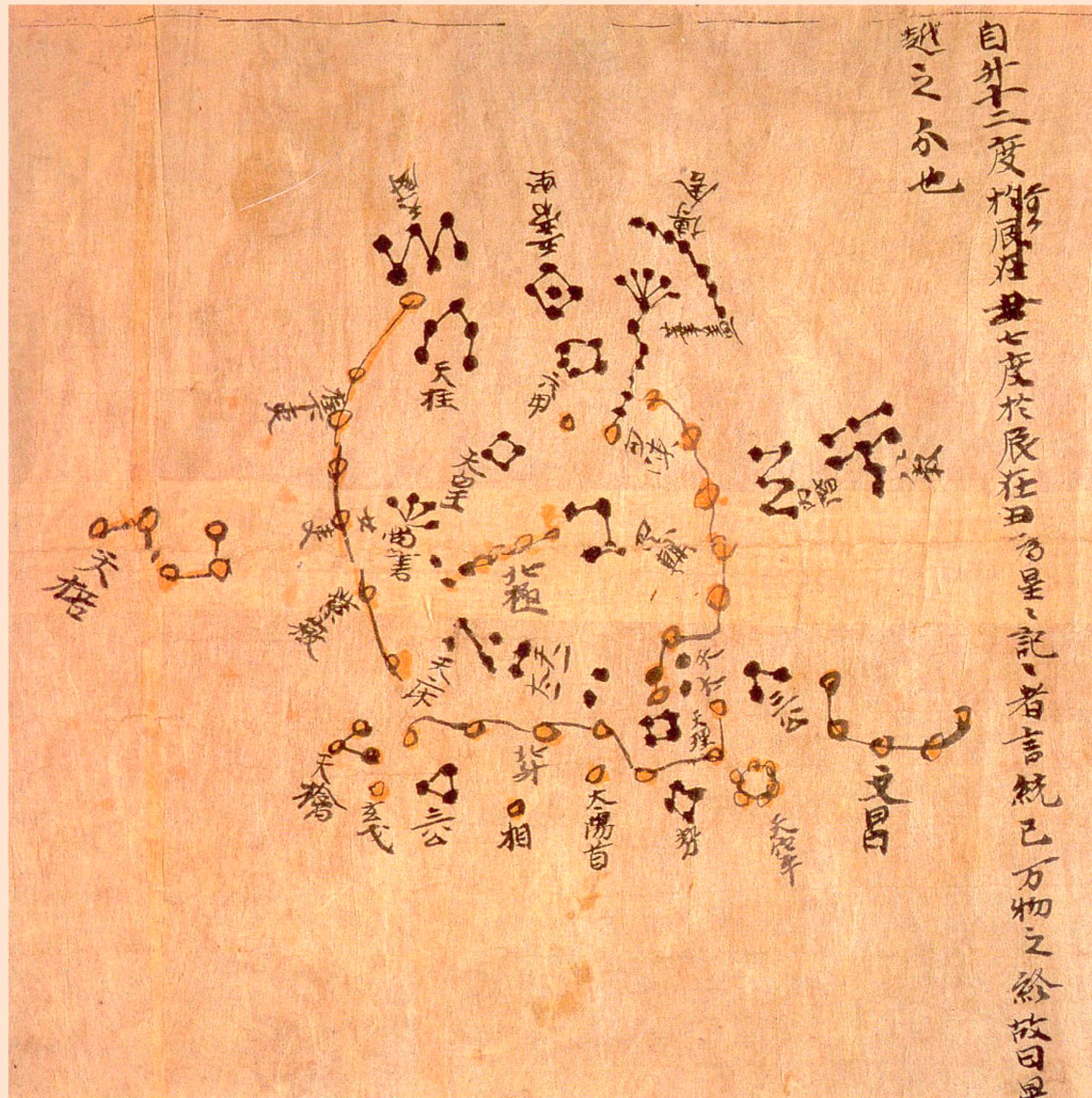


(A reprint of a map showing labeled seas, roads, and small symbols for villages and cities)

705-1000 CE

Star Chart

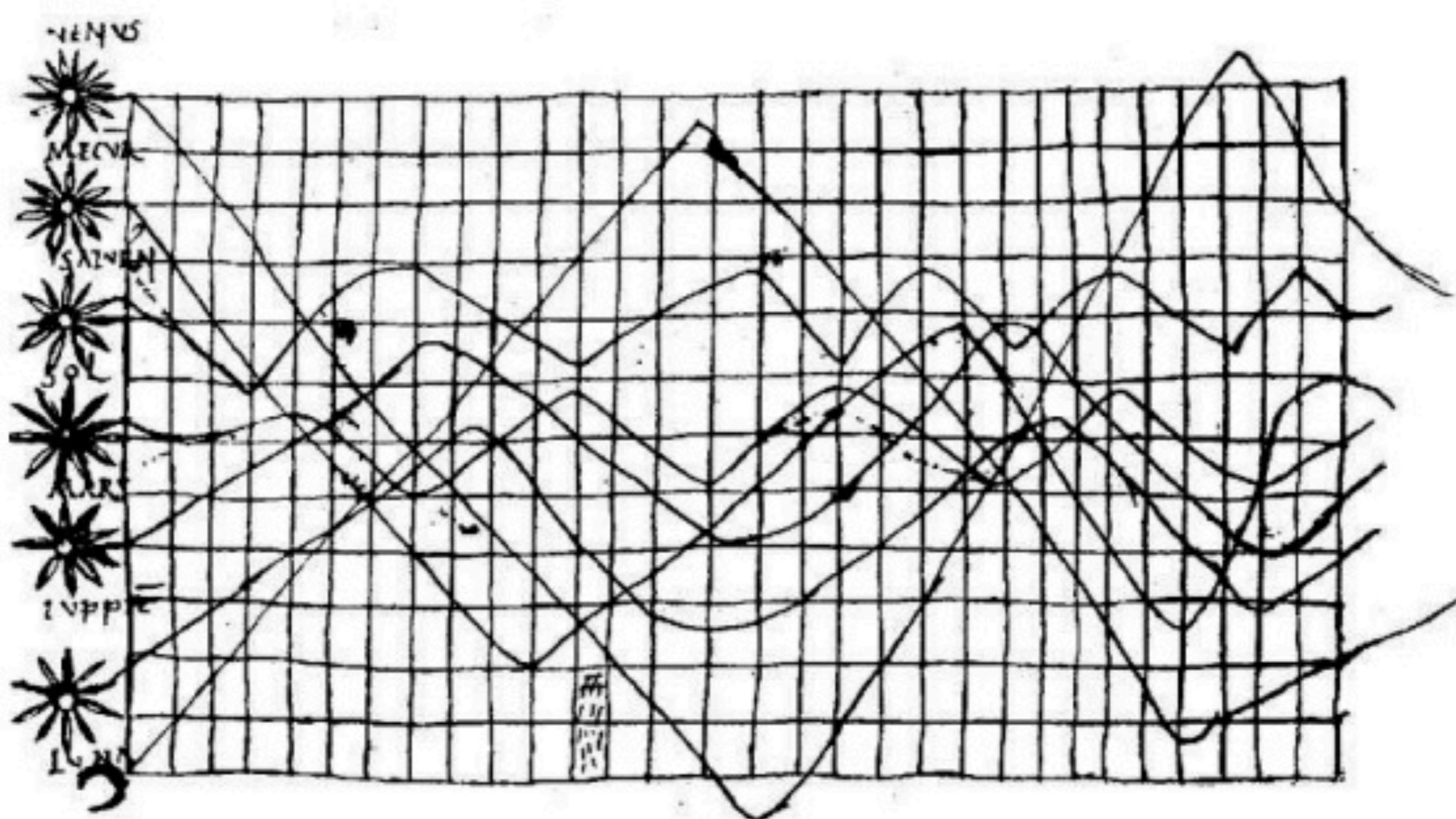
Tang Dy.



(Chinese calligraphy annotating diagrams of stellar constellations)

~950 CE

Macrobius - Planetary Motion



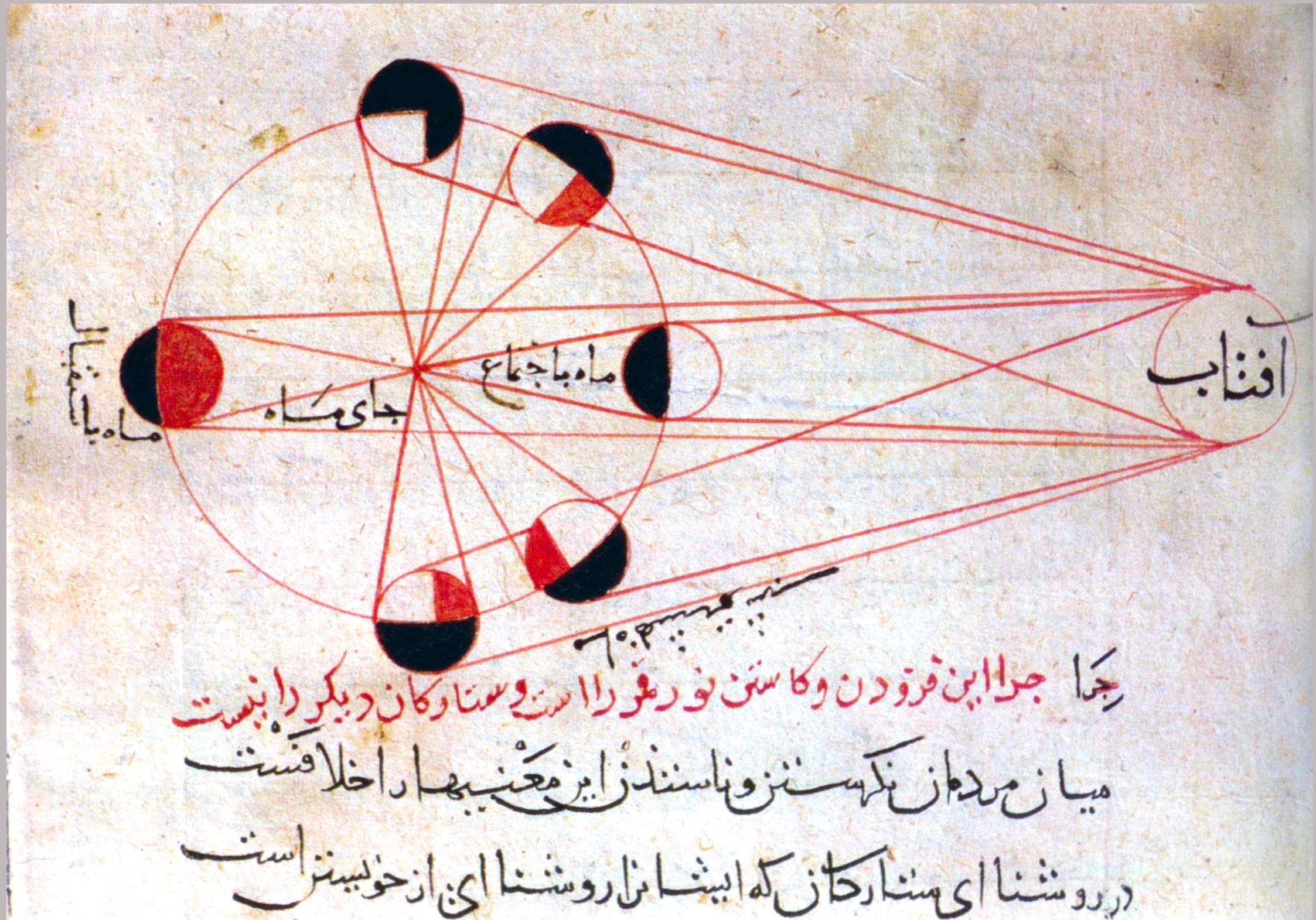
(Star markers with Latin planet names connect to curved lines on a grid)

Img credit
Funkhouser, 1936



~1000 CE

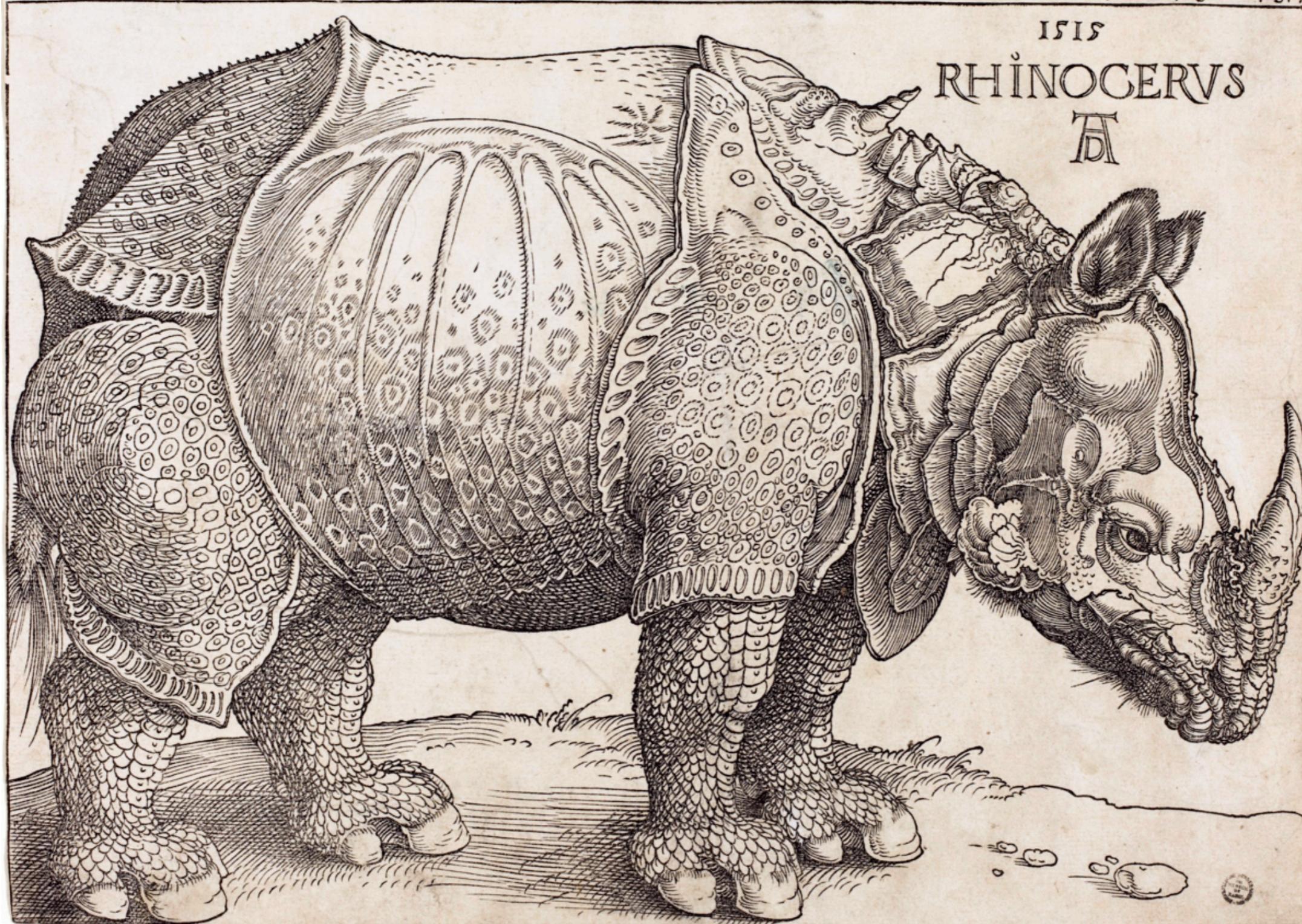
Al Biruni - Lunar Phases



(Arabic script annotating a diagram of lunar phases marked in white, red, and black)

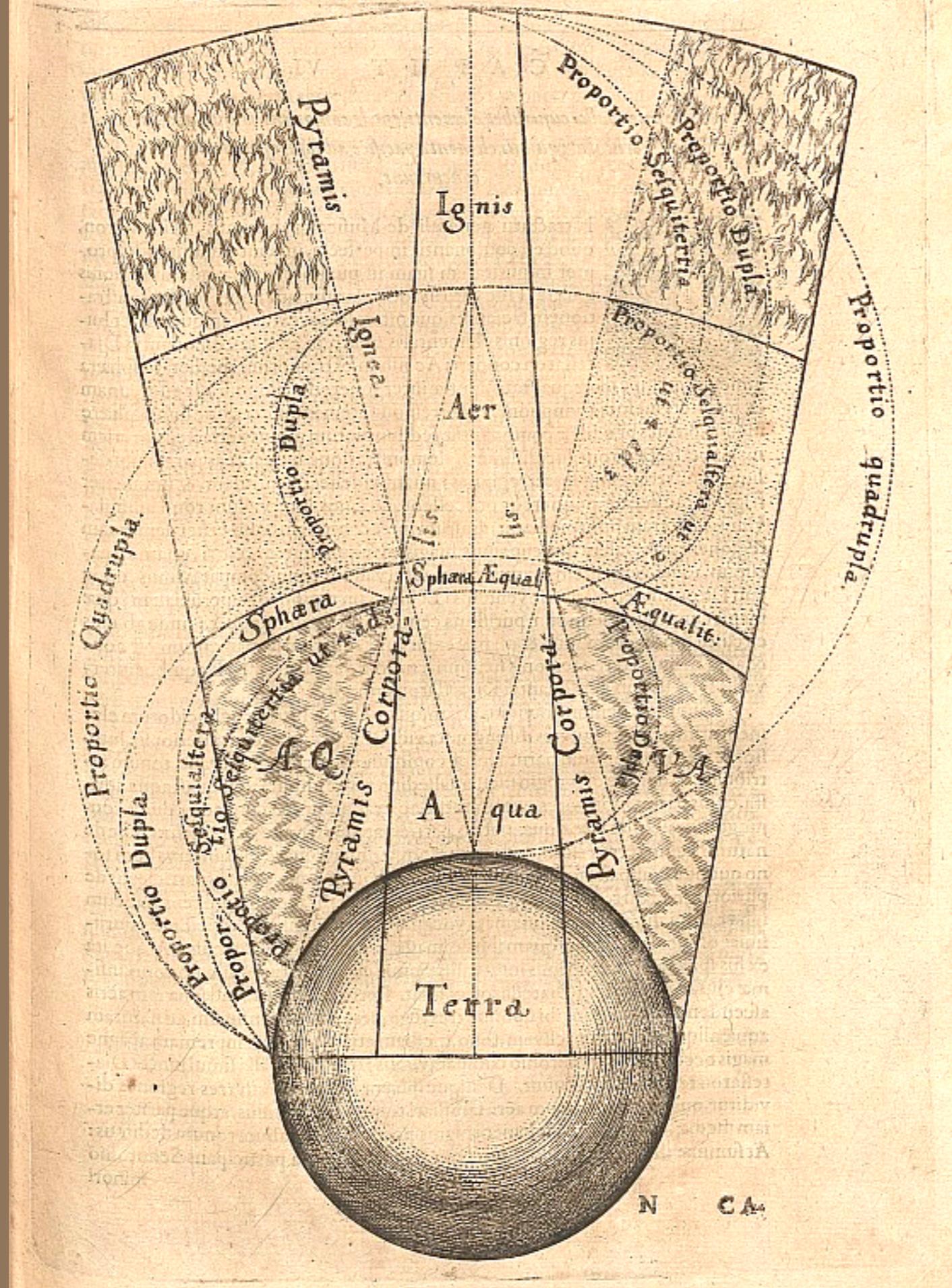


Nach Christus gepint. 1515. Jar. Adi. 1. May. Hat man dem grossmächtigen Künig von Portugall Emanuell gen Lysabona pracht auf India/ein sollich lebendig Thier. Das nennen sie Rhinocerus. Das ist hye mit aller seiner geslalt Abscondertset. Es hat ein farb wie ein gespreckete Schildkrot. Und ist vō dicke[n] Schalen überlegt fast fest. Und ist in der groß als der Helfsandt Aber nyderrechtiger von paynen/vnd fast wehafftig. Es hat ein scharff stark Horn vom auff der nasen/Das begyndt es albeg zu wezen wo es bey staynen ist. Das dosig Thier ist des Helffangz todt seyndt. Der Helfsandt fürcht es fast vbel/dann wo es In ankumbt/so lauftt Im das Thier mit dem Kopff zwischen dye fordern payn/vnd reyst den Helfsandt vnden am pauch auss vñ erwürgt In/des mag er sich nit erweren. Dann das Thier ist also gewapent/das Im der Helfsandt nichts kan thün. Sie sagen auch das der Rhinocerus Schnell/graydig vnd Listig sey.



(An engraving with text and a depiction of a rhinoceros in black and white ink)

1617 CE

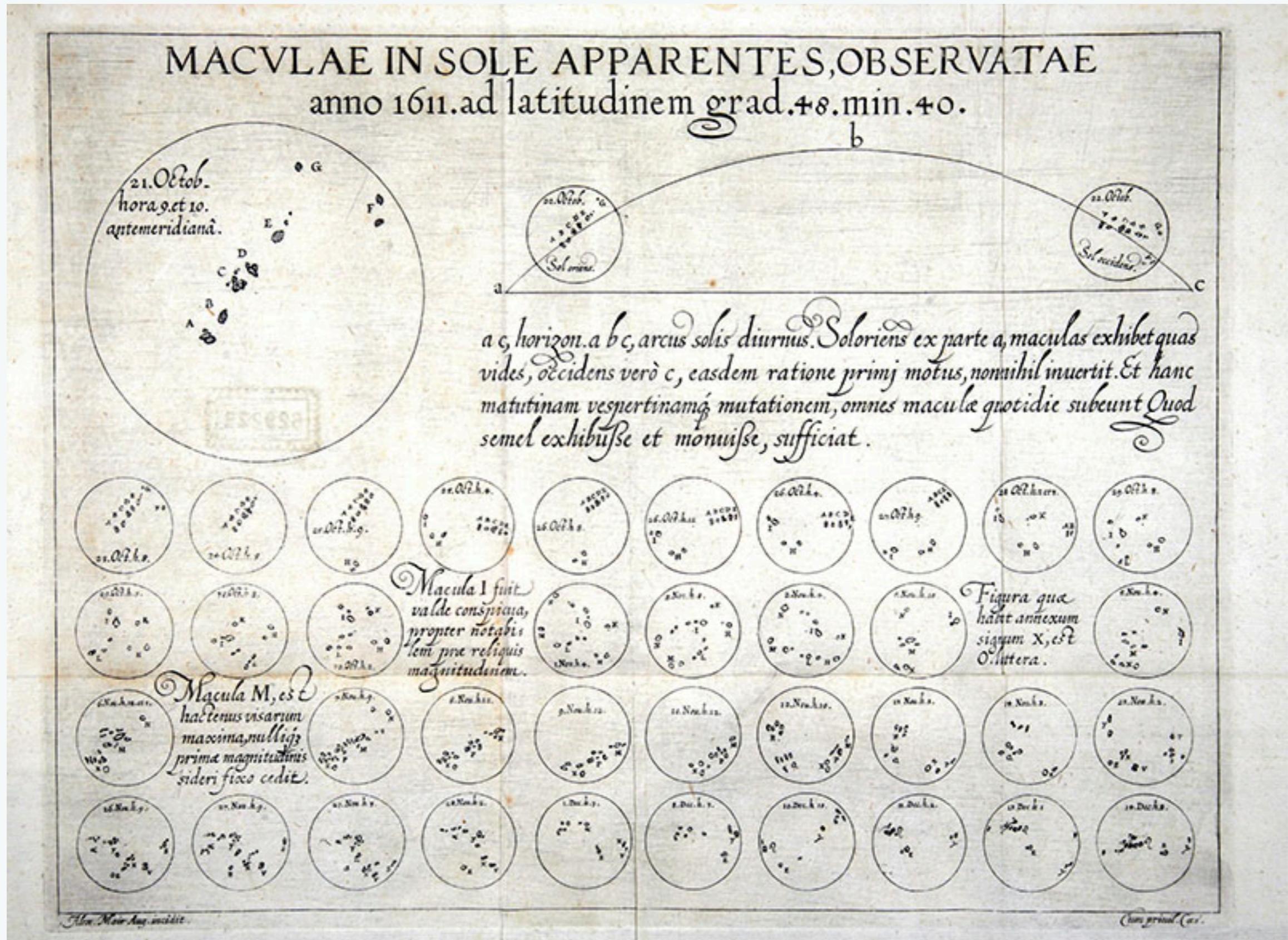


Fludd - Classical Elements

(A black and white print depicting fire, air, water, and earth (top-bottom))

1626 CE

Schiener - Sunspots over time



(Latin text annotates a series of circular diagrams showing sunspot positions at given dates)



1644 CE

Longitude Estimates

TOLEDO.

GRADOS DE LA LONGITUD.

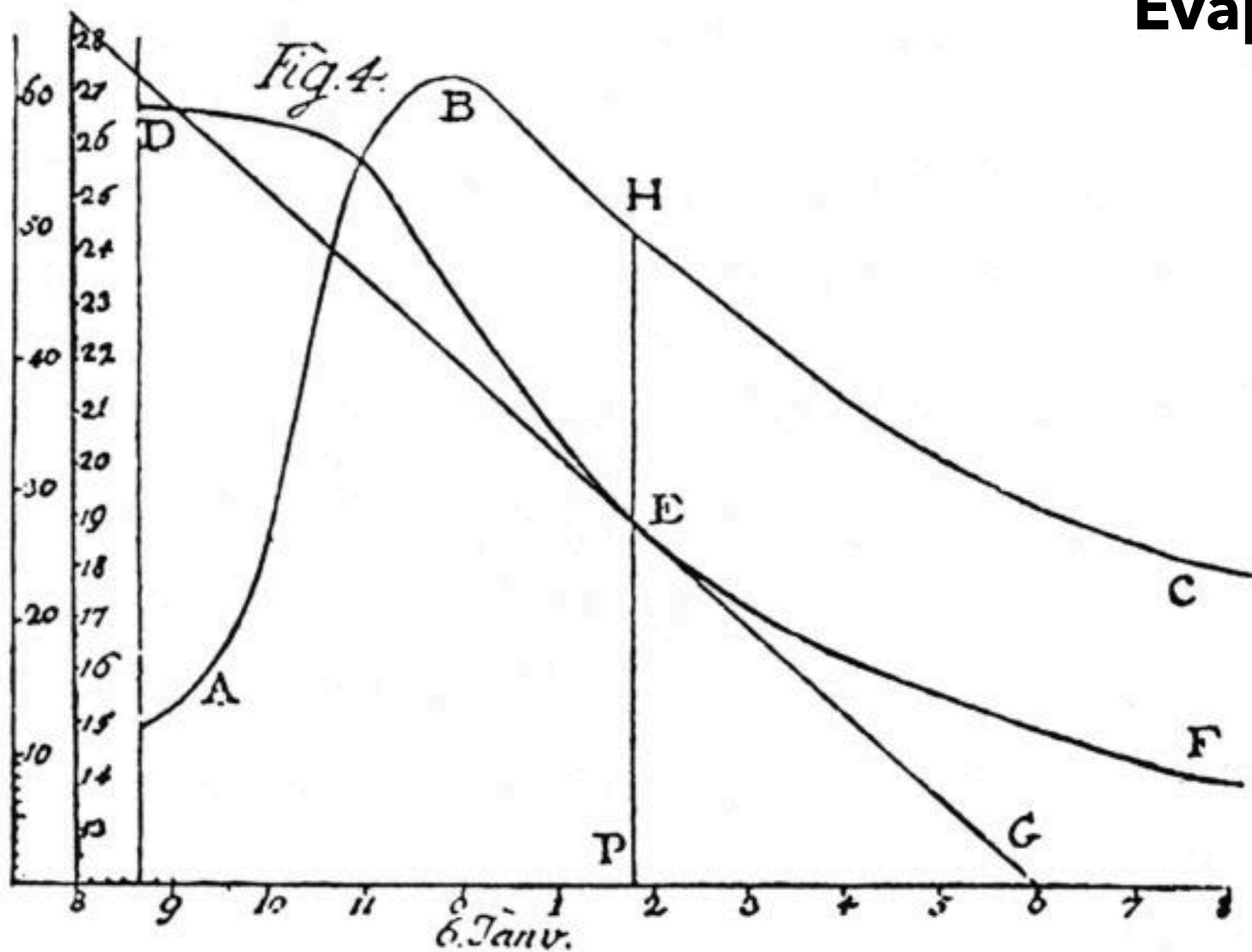


(City names mark places on a number line going from 0 (left) to 30 (right). An arrow marks Van Langren's estimate; a Google map view below overlays VL's original chart.)



1765 CE

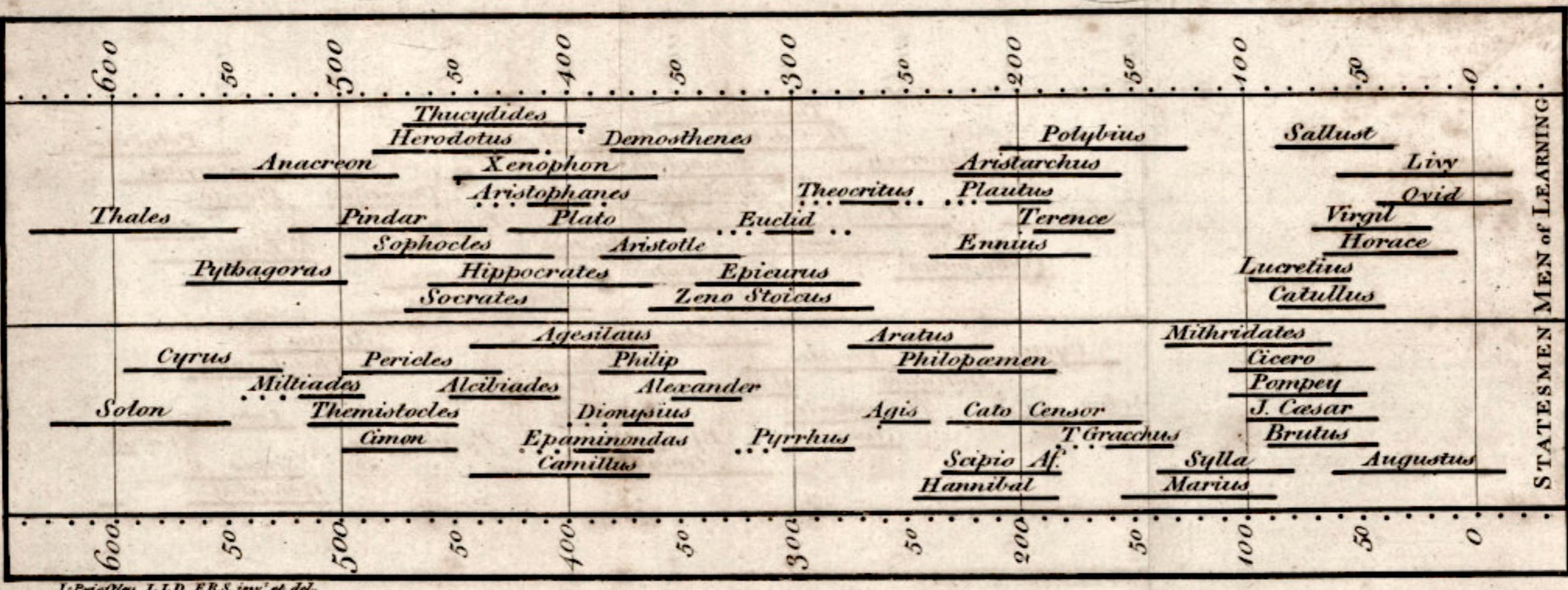
Lambert
Rate of Water
Evaporation



(Curved lines above an X and Y scale)

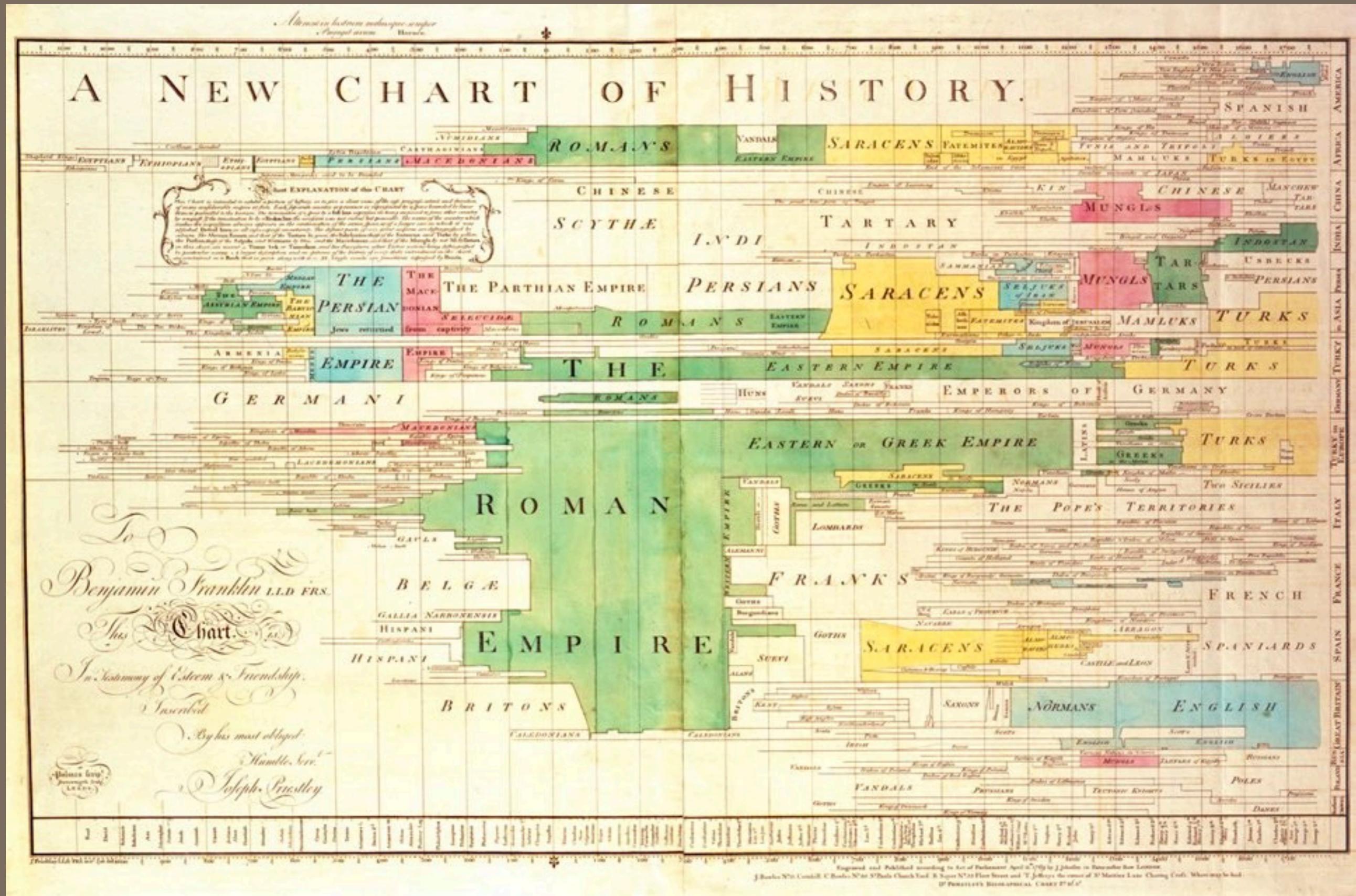


A Specimen of a Chart of Biography.



(Year markers on indicate starting and ending years for lines corresponding to famous individuals)



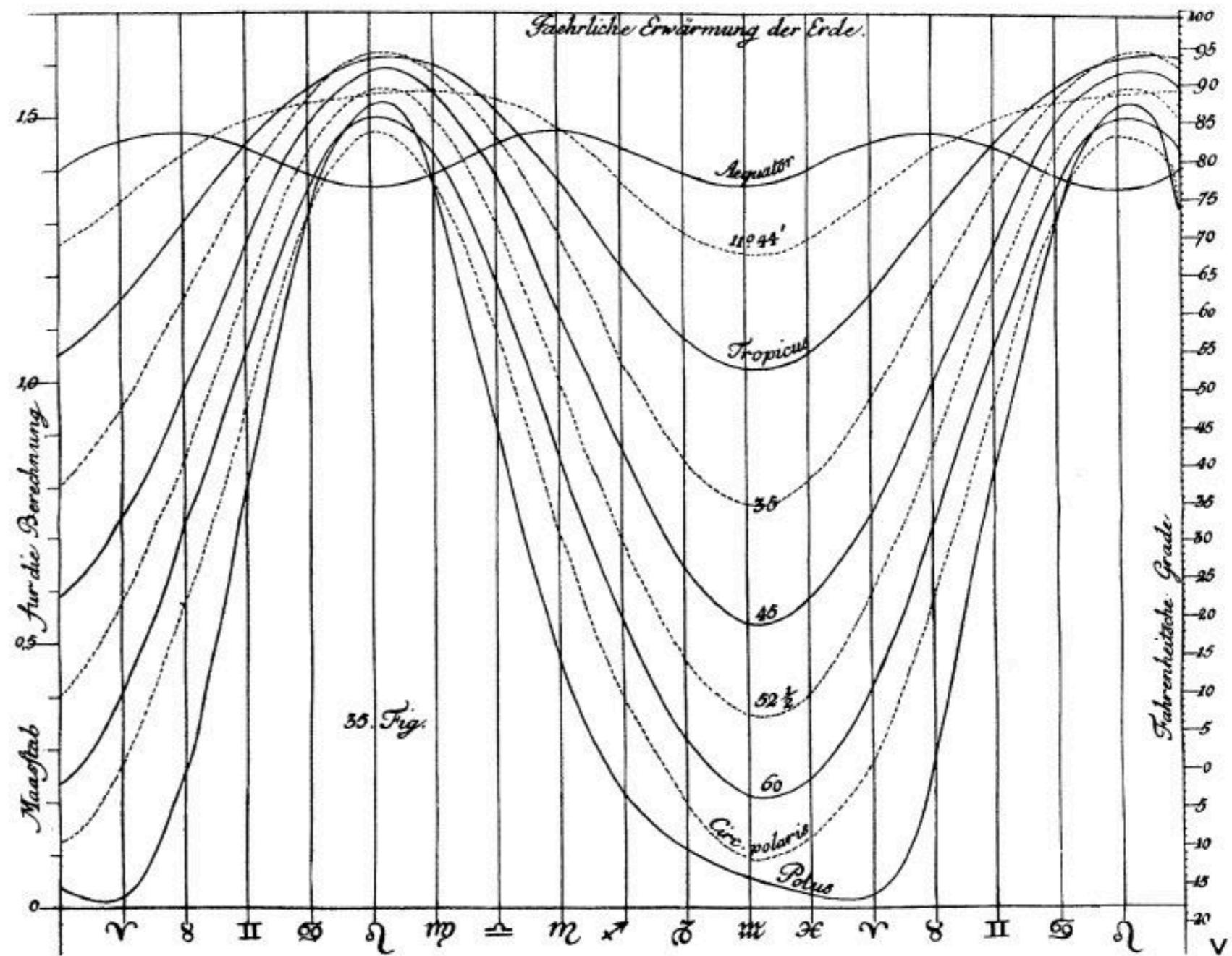


1779 CE

Lambert

Soil Temperatures

	11	12	13	14	15	16	17	18	19	20	21	22	23
Jan.						1		4	8	9	4	4	7
Feb.													9
Mar.													8
Apr.													7
May													6
Jun.													5
Jul.													4
Aug.													3
Sept.													2
Oct.													1
Nov.													0
Dec.													-1
1735													-2
1736													-3



(A data table and associated printed line chart. Symbols indicate months)



1759-1823 CE

William Playfair

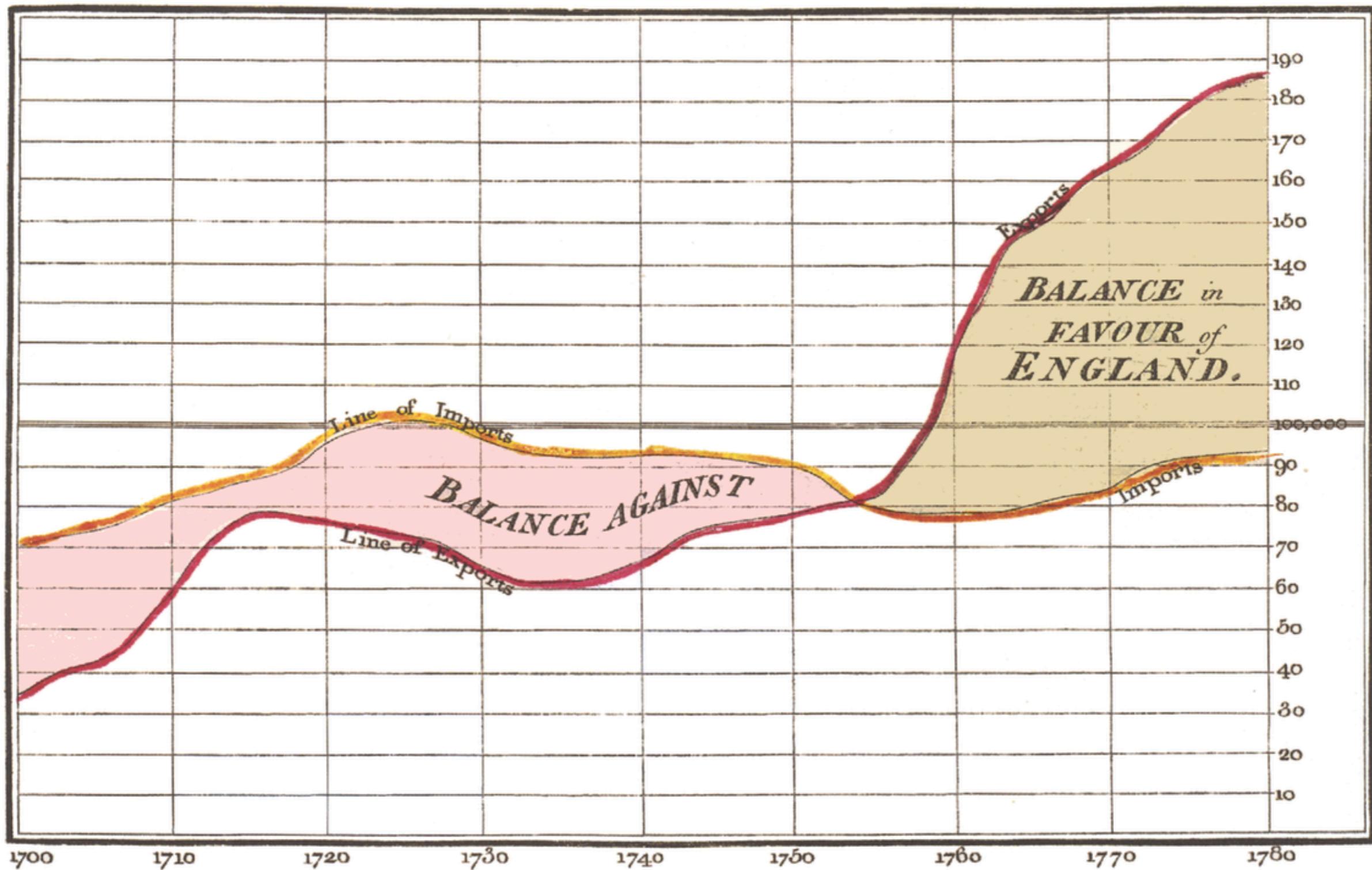


(An etching of Scottish polymath William Playfair)

1786 CE

Playfair

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



The Bottom line is divided into Years, the Right hand line into £10,000 each.
Published as the Act directs, 1st May 1786, by Wm. Playfair
Neale sculp^t 352, Strand, London.

(Red and yellow regions on a line chart highlight trade imbalance)



1801 CE

Playfair

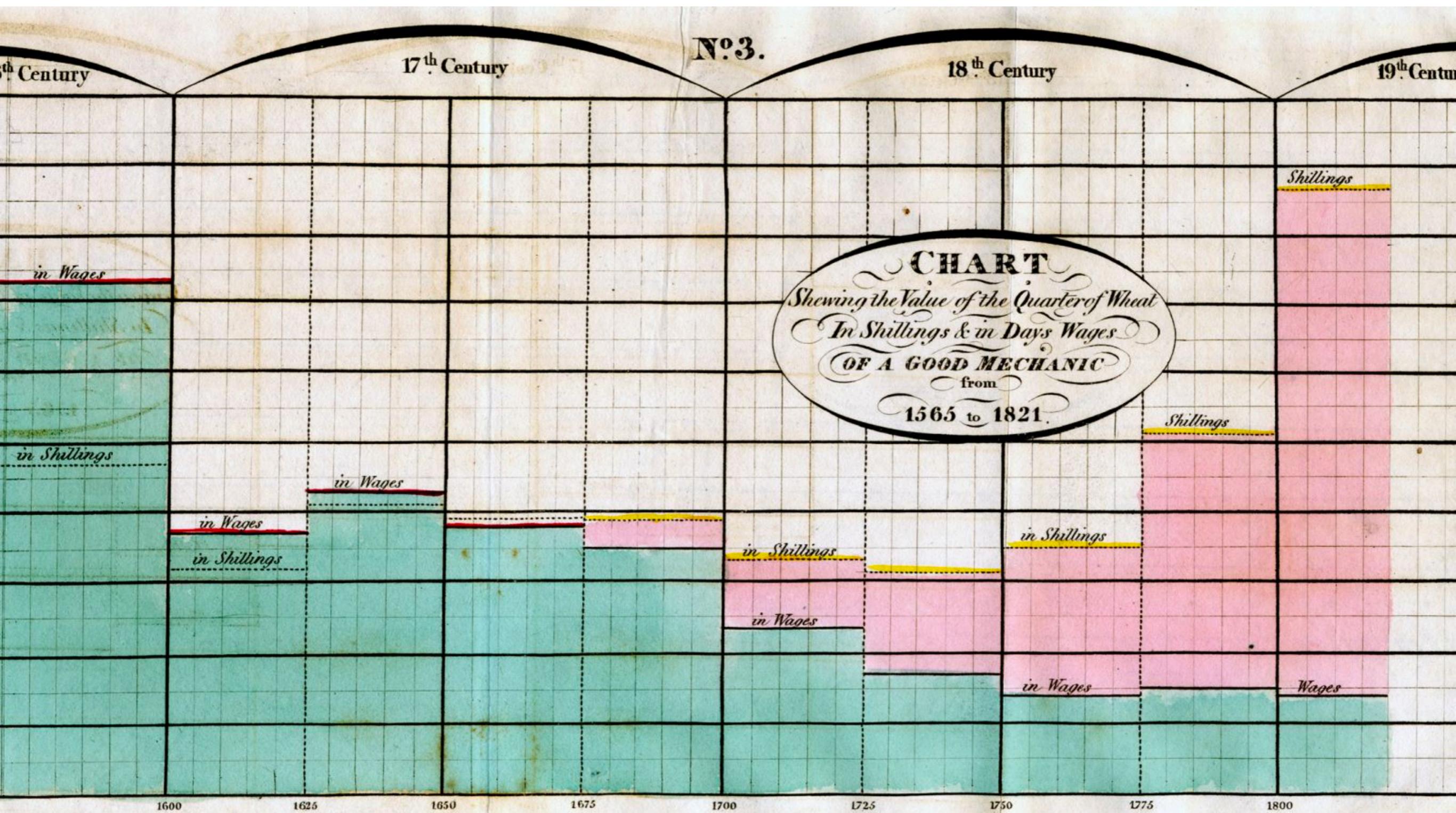


(Multicolored circles connected to bars show land area of empires versus population and tax income)



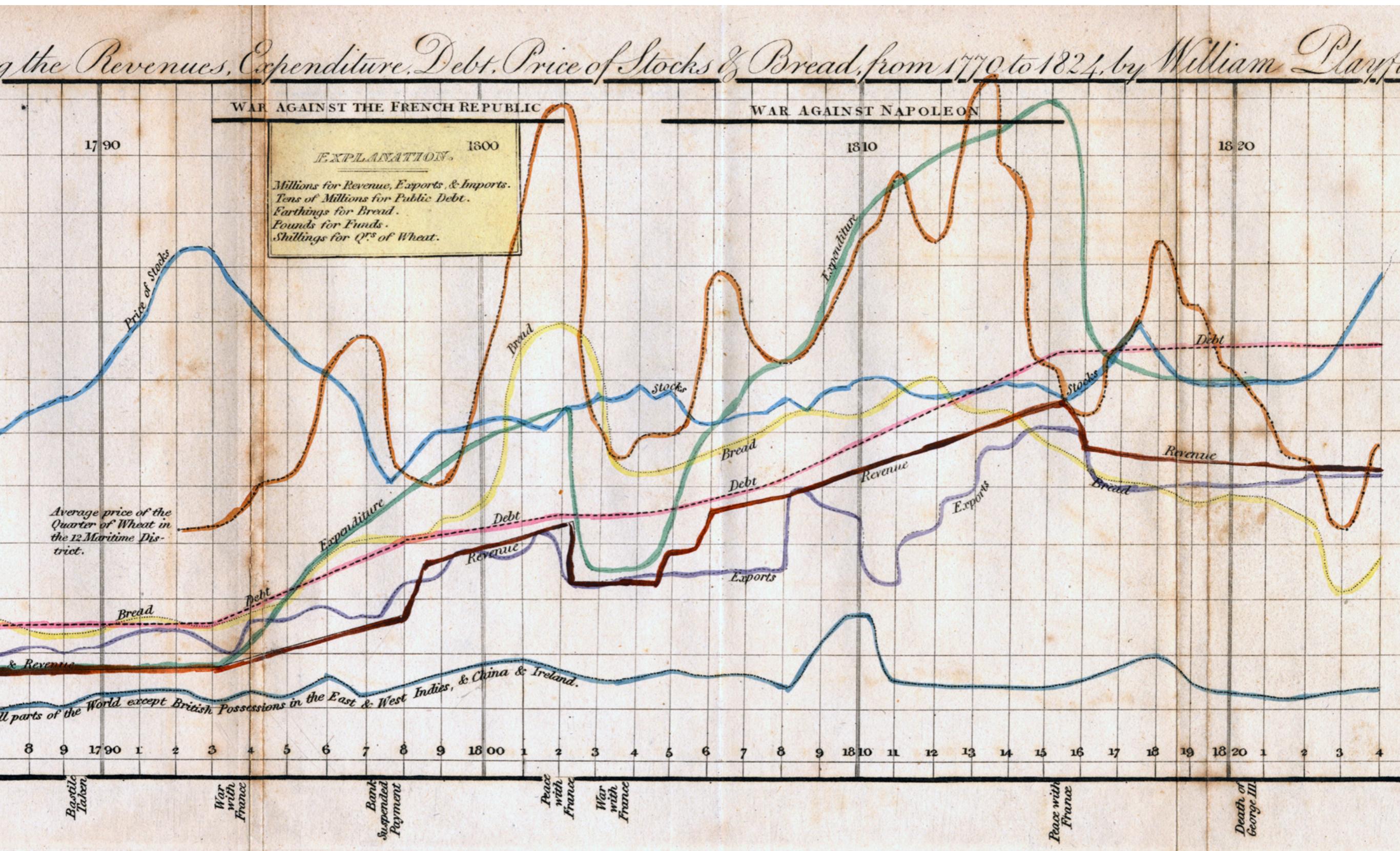
1821 CE

Playfair



(Stacked bars in blue and pink show the cost of wheat at various times [wages blue, shillings pink])

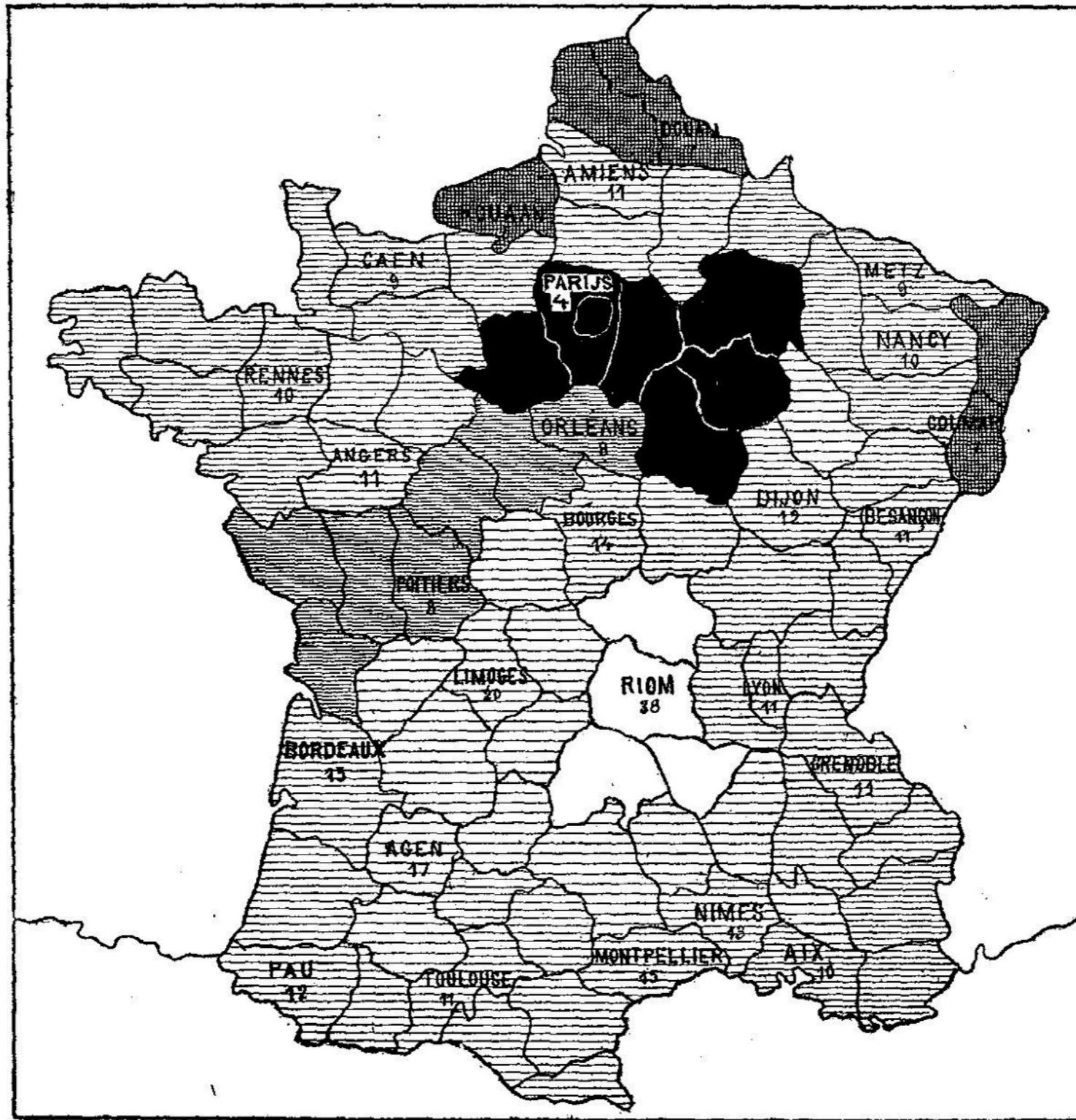




(Multicolored lines chart the price of different commodities during war and peace [x axis time])

1826 CE

Dupin



(A choropleth map of France; choropleth = "many regions")



1820-1910 CE

Florence Nightingale

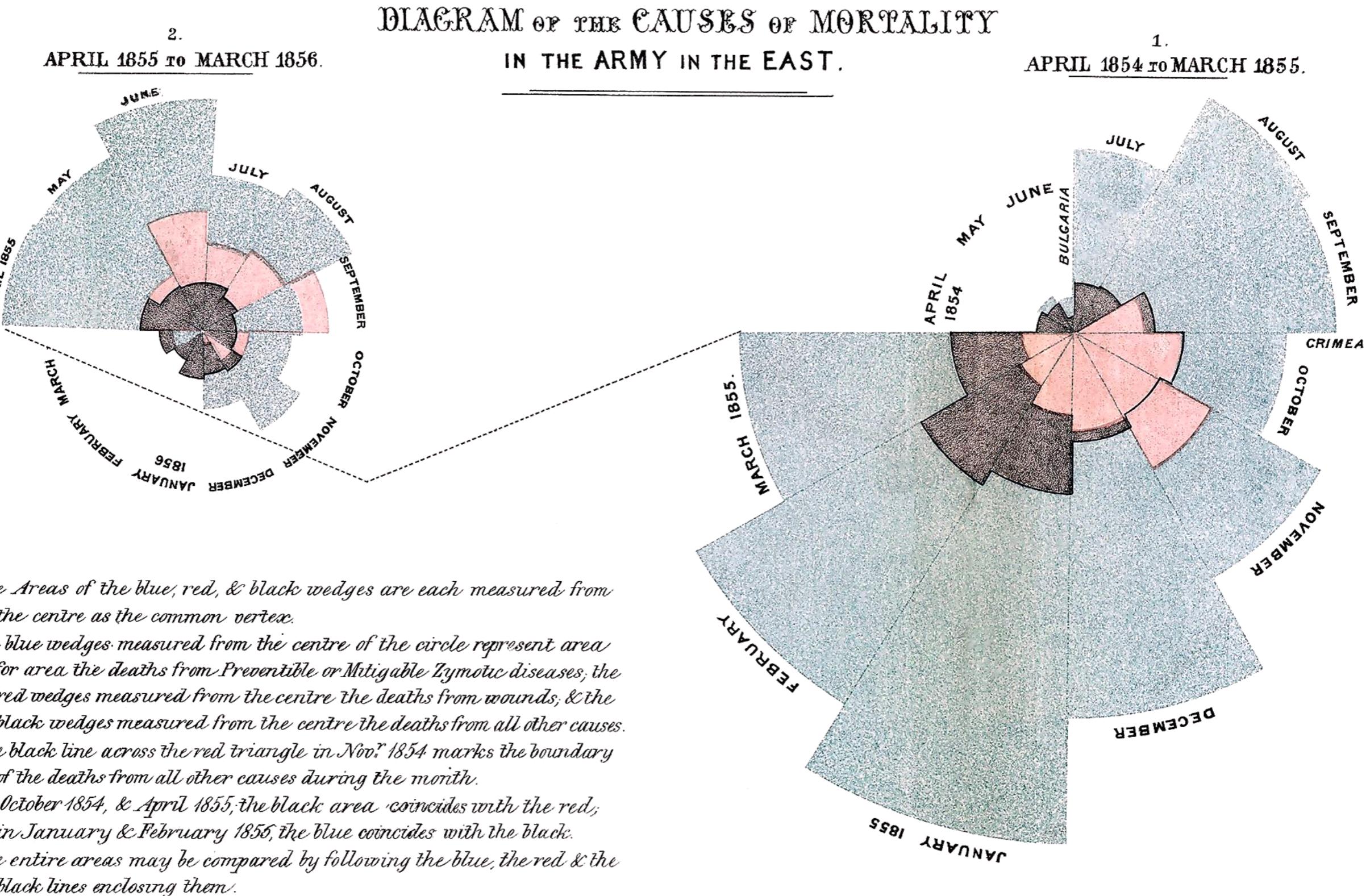


(A photograph of British statistician, nurse and medical reform advocate Florence Nightingale)



1856 CE

Nightingale



(A "rose diagram" depicting different amounts (length) of military casualties (color) over time (angle)) (Crimean war)



1855 CE

John Snow



(Circle markers on a map indicate deaths due to cholera, Xs mark water pumps)

(Redrawn by Tufte)



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 Séguir, 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 Mobilow et qui rejoignirent vers Orscha et Witebsk, avaient toujours marché avec l'armée.

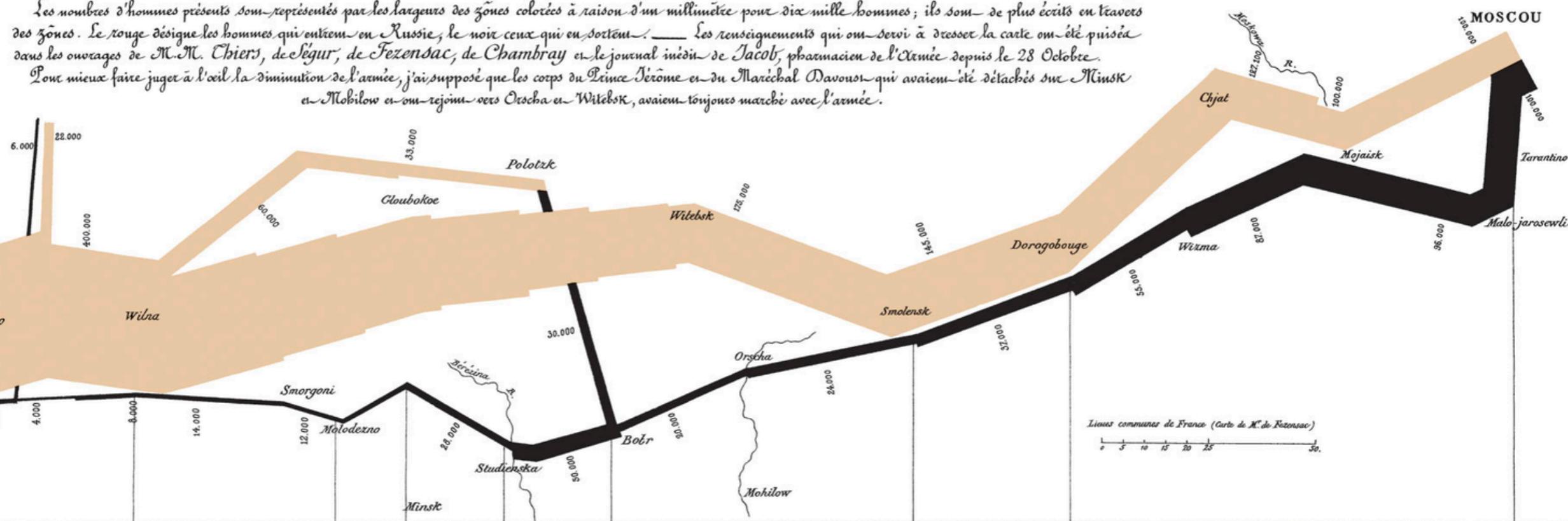
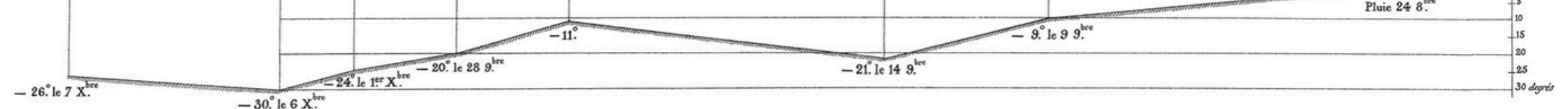


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

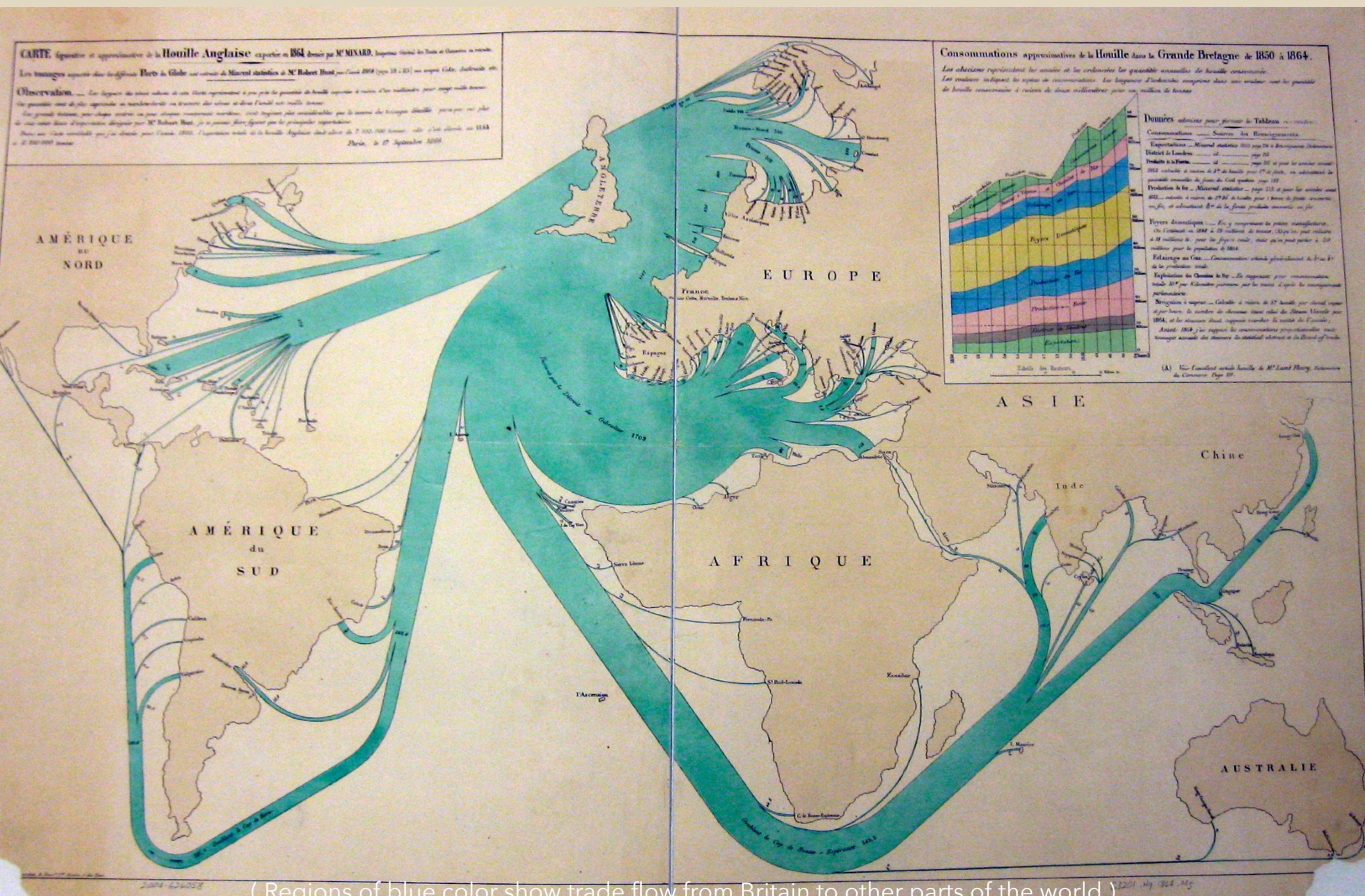


Autog. par Régnier, 8. Pas. S^e Marie St^e G^e à Paris.

Imp. Lith. Régnier et Dourdet.

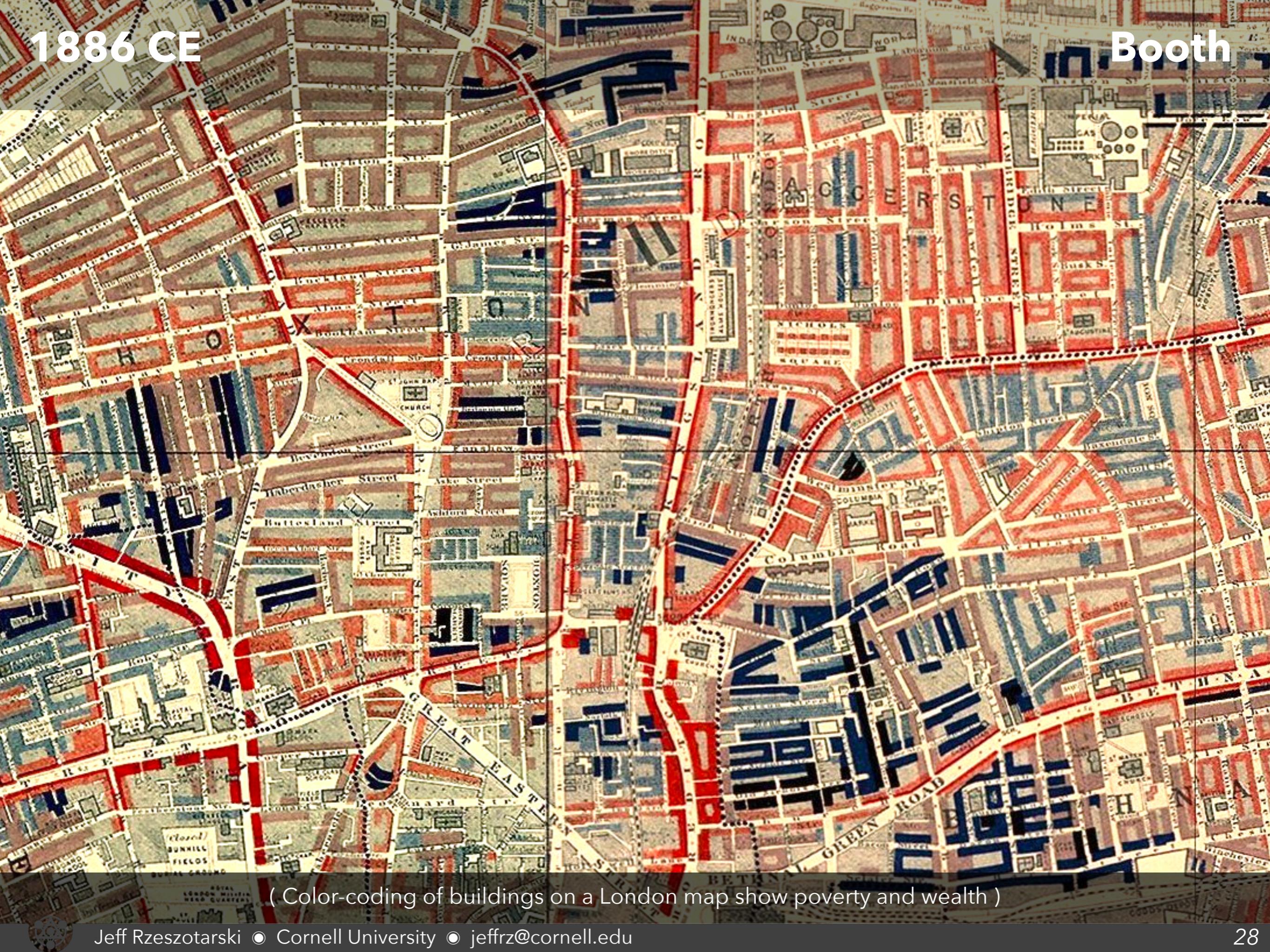
(Colored lines show the march to Russia (tan) and return (black) as army size (width) diminishes. A line chart shows parallel temperature decreases over time)





1886 CE

Booth

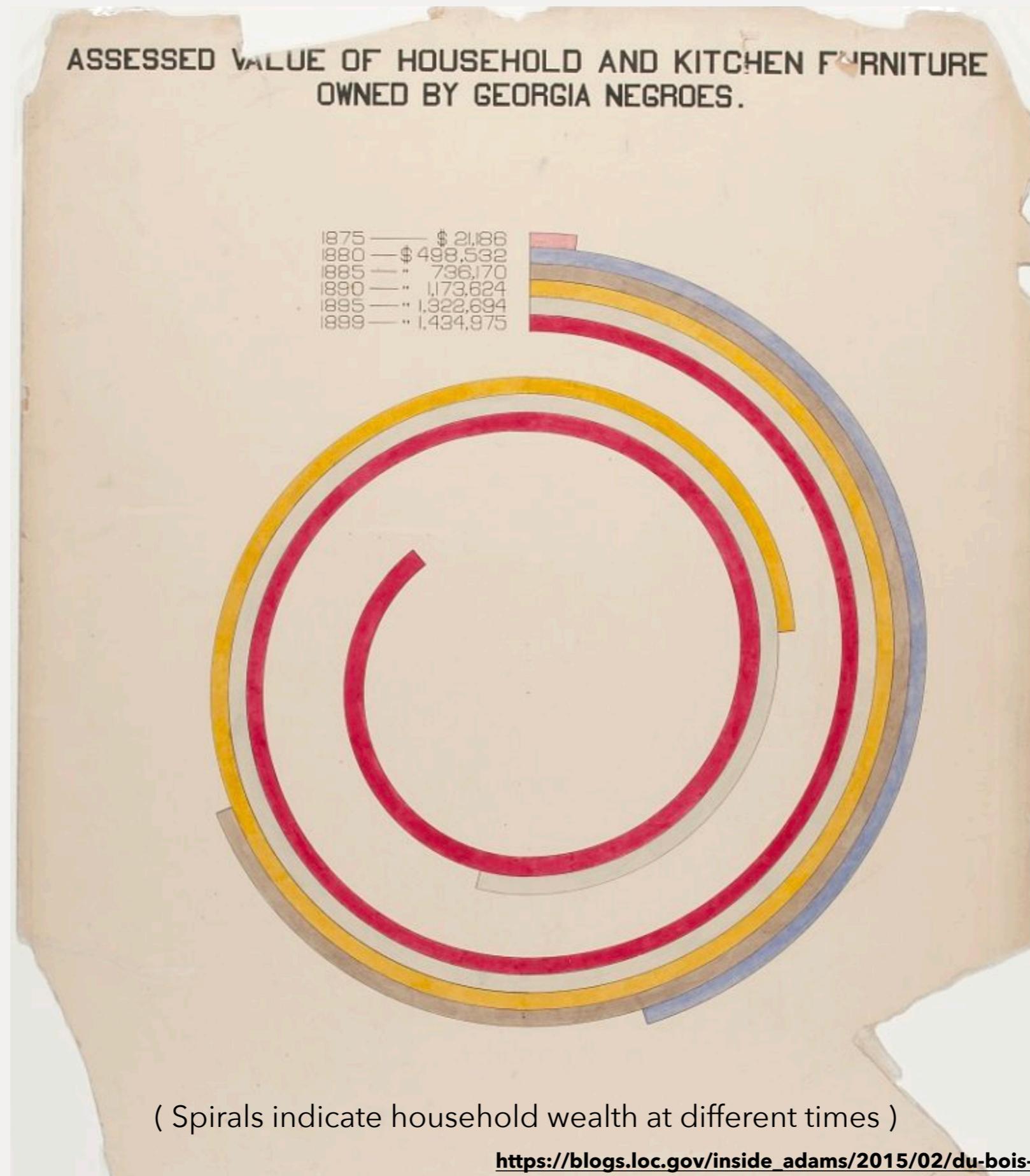


(Color-coding of buildings on a London map show poverty and wealth)



1900 CE

W.E.B. DuBois - Paris Exposition



Foundation of modern statistics

Applying and popularizing viz in the mainstream,
but not developing new forms of visualization

1962 CE

John Tukey

"The Future of Data Analysis"



(A photograph of John Tukey in his laboratory)



1962 CE

John Tukey

Stats Methods

Computers

Lots of Data

Quantitative Decision-making



Stats Methods



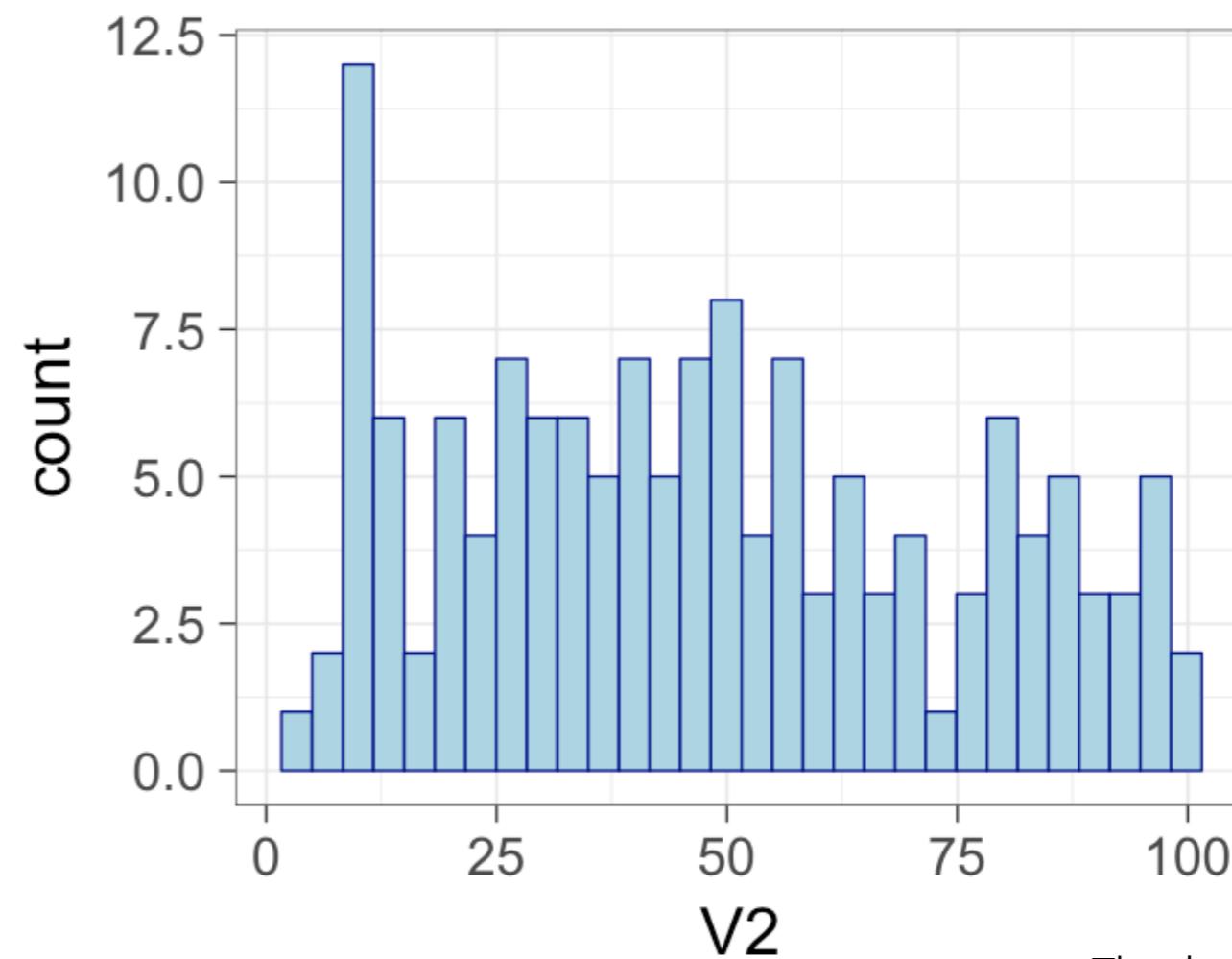
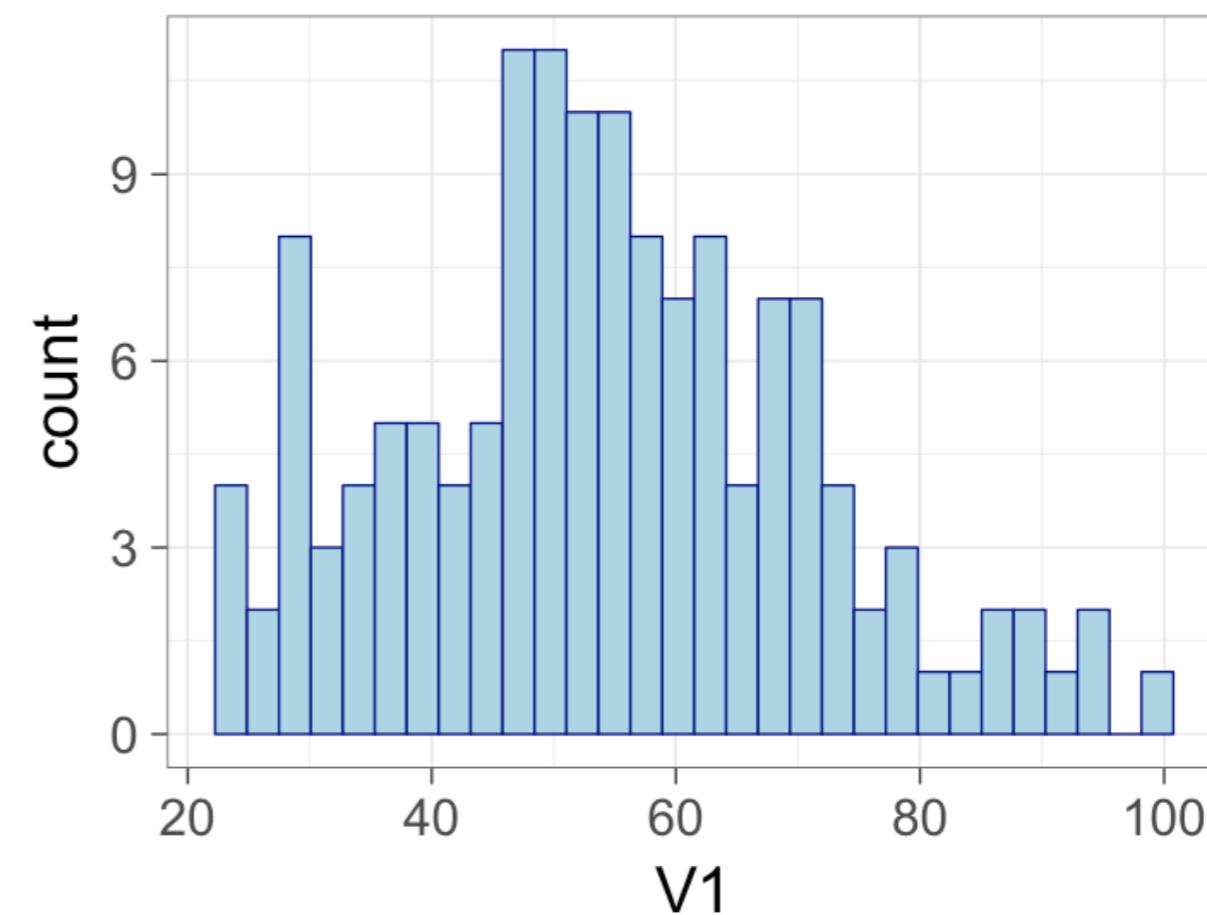
Computers — DANGER — Lots of Data



Quantitative Decision-making

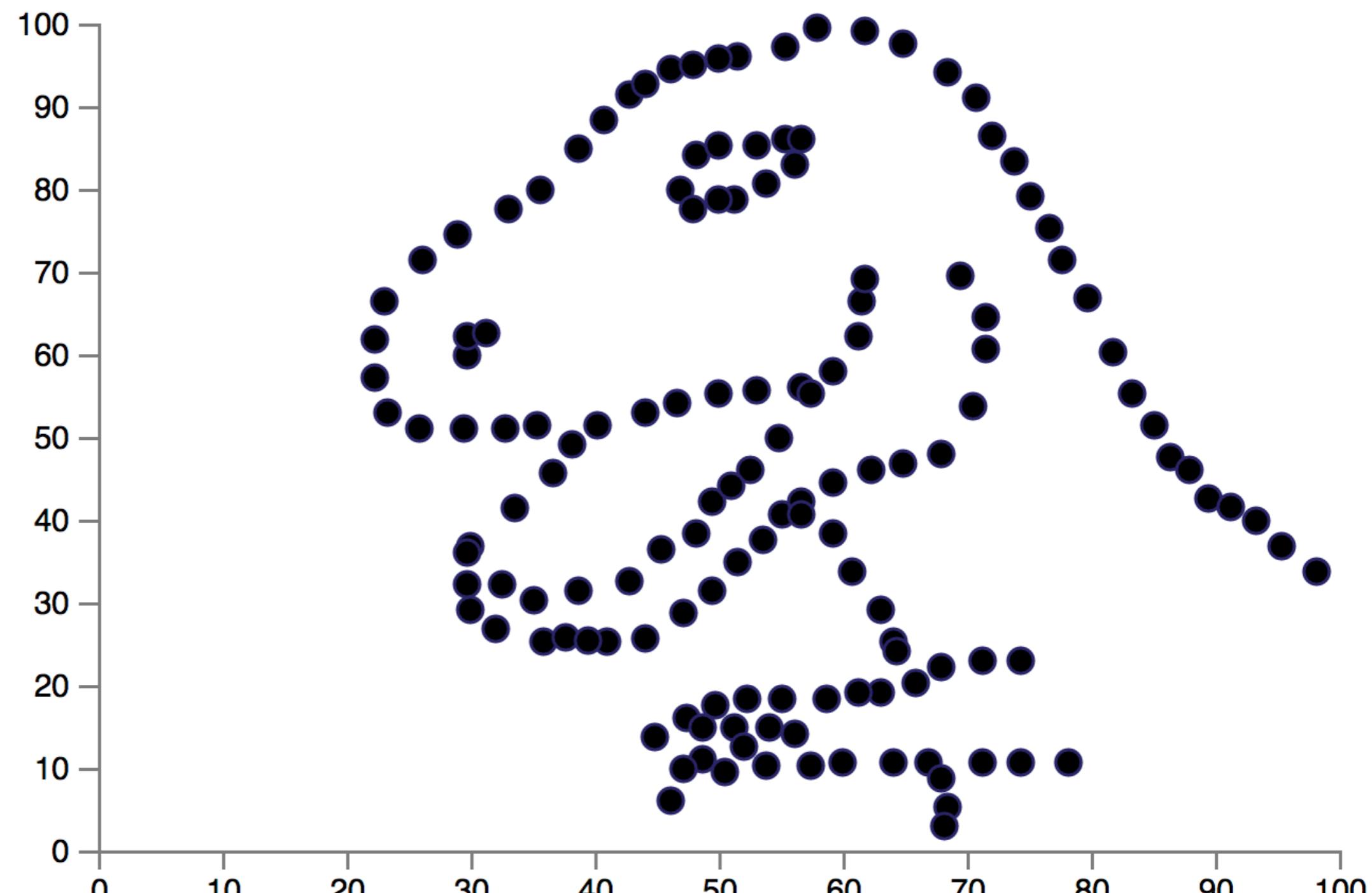
V1	V2
55.3846	97.1795
51.5385	96.0256
46.1538	94.4872
42.8205	91.4103
40.7692	88.3333
38.7179	84.8718
35.641	79.8718
33.0769	77.5641
28.9744	74.4872
26.1538	71.4103
23.0769	66.4103
22.3077	61.7949
22.3077	57.1795
23.3333	52.9487
25.8974	51.0256
29.4872	51.0256
32.8205	51.0256
35.3846	51.4103
40.2564	51.4103
44.1026	52.9487
46.6667	54.1026
50	55.2564
53.0769	55.641
56.6667	56.0256
59.2308	57.9487
61.2821	62.1795

V1	V2
55.3846	97.1795
51.5385	96.0256
46.1538	94.4872
42.8205	91.4103
40.7692	88.3333
38.7179	84.8718
35.641	79.8718
33.0769	77.5641
28.9744	74.4872
26.1538	71.4103
23.0769	66.4103
22.3077	61.7949
22.3077	57.1795
23.3333	52.9487
25.8974	51.0256
29.4872	51.0256
32.8205	51.0256
35.3846	51.4103
40.2564	51.4103
44.1026	52.9487
46.6667	54.1026
50	55.2564
53.0769	55.641
56.6667	56.0256
59.2308	57.9487
61.2821	62.1795



Thanks, Alberto Cairo

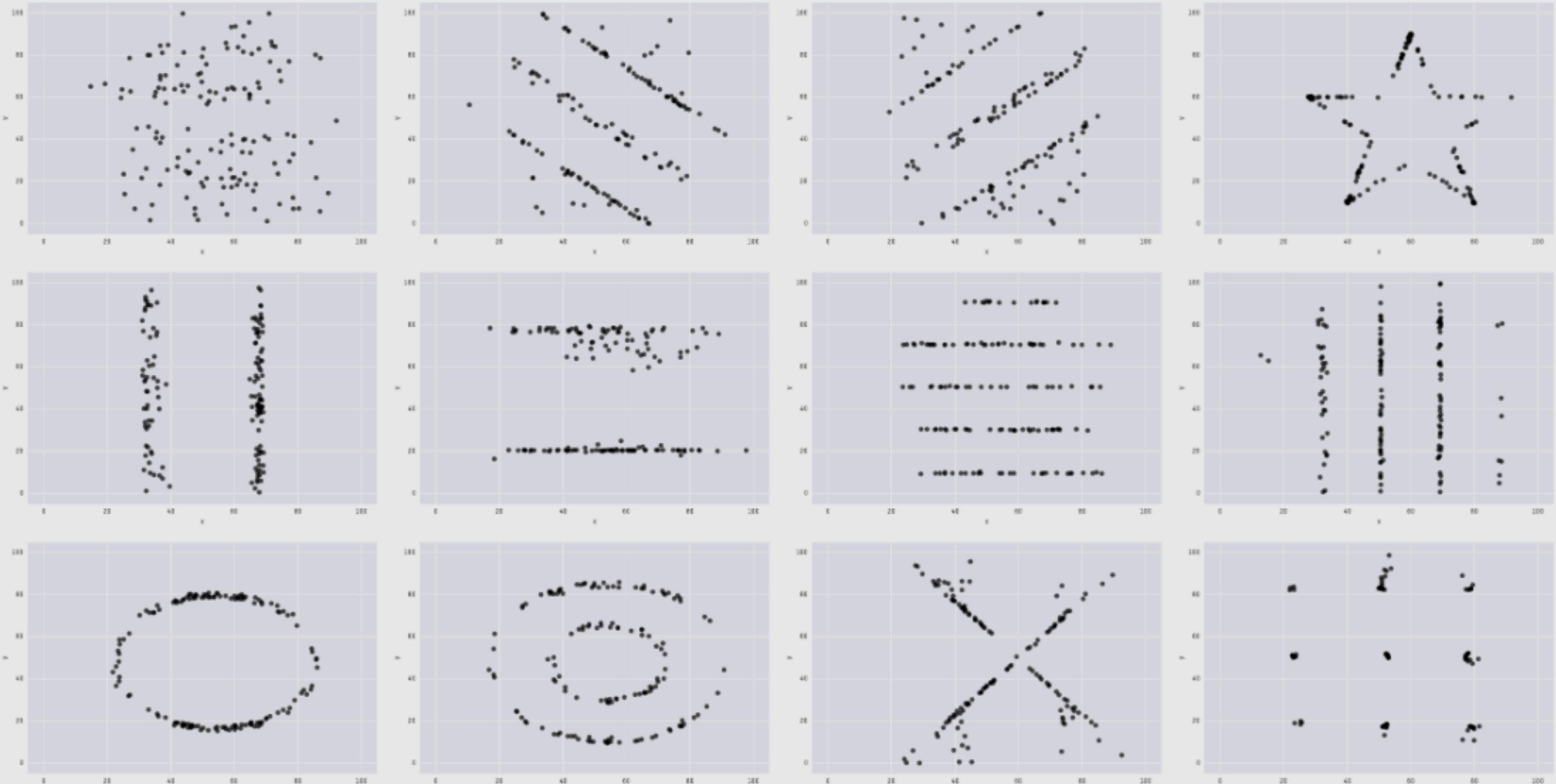
V1	V2
55.3846	97.1795
51.5385	96.0256
46.1538	94.4872
42.8205	91.4103
40.7692	88.3333
38.7179	84.8718
35.641	79.8718
33.0769	77.5641
28.9744	74.4872
26.1538	71.4103
23.0769	66.4103
22.3077	61.7949
22.3077	57.1795
23.3333	52.9487
25.8974	51.0256
29.4872	51.0256
32.8205	51.0256
35.3846	51.4103
40.2564	51.4103
44.1026	52.9487
46.6667	54.1026
50	55.2564
53.0769	55.641
56.6667	56.0256
59.2308	57.9487
61.2821	62.1795



Thanks, Alberto Cairo



X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06



Very cool demo: <https://www.autodeskresearch.com/publications/samestats>



I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

M = 9 and 7.5

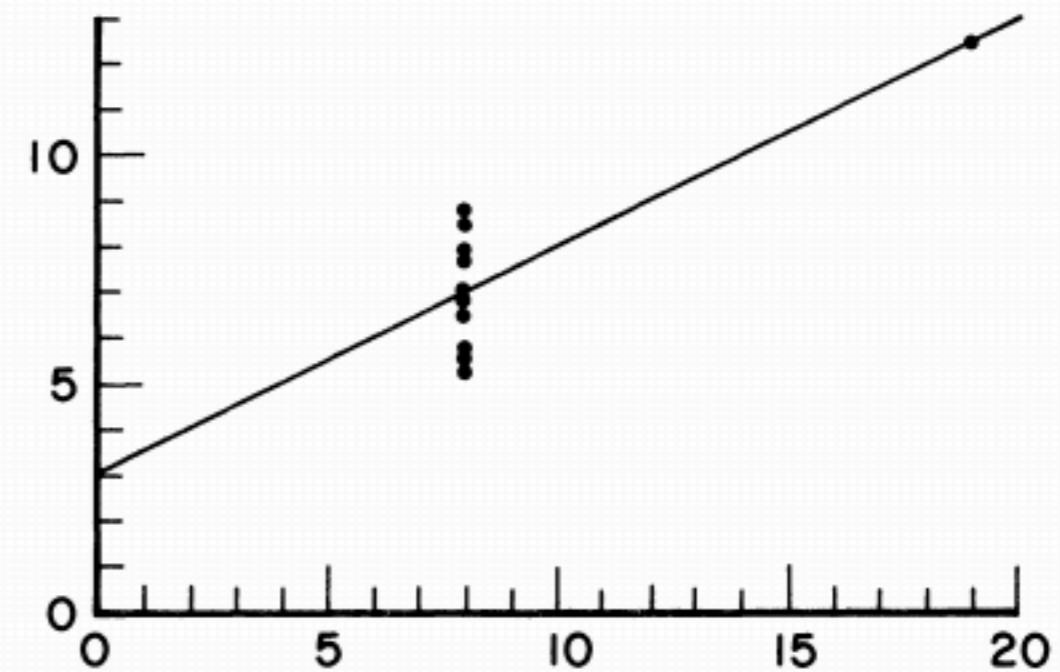
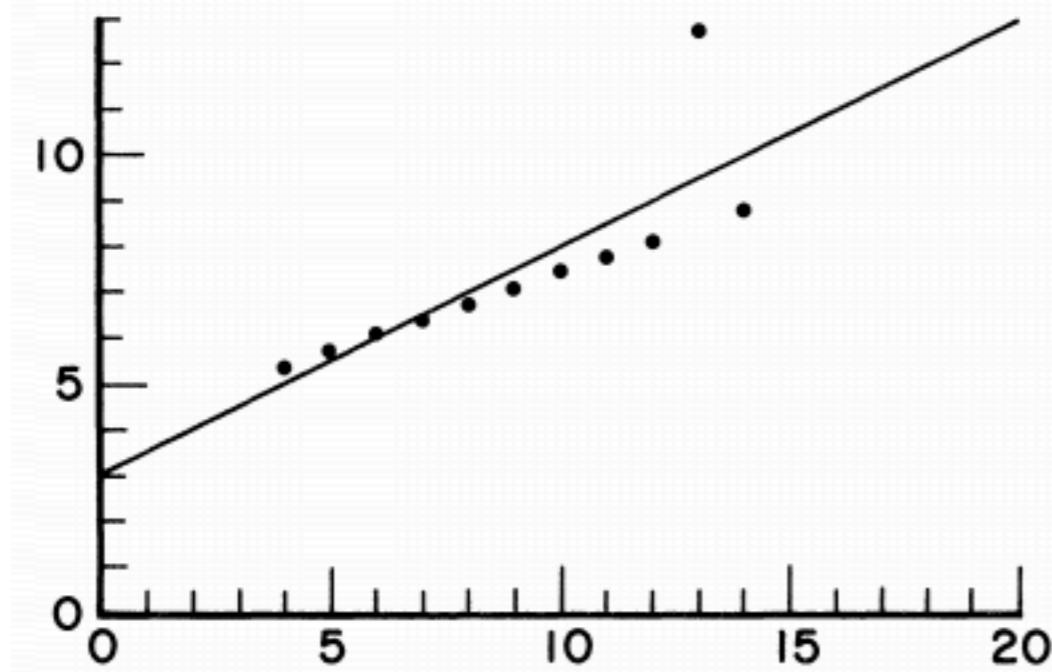
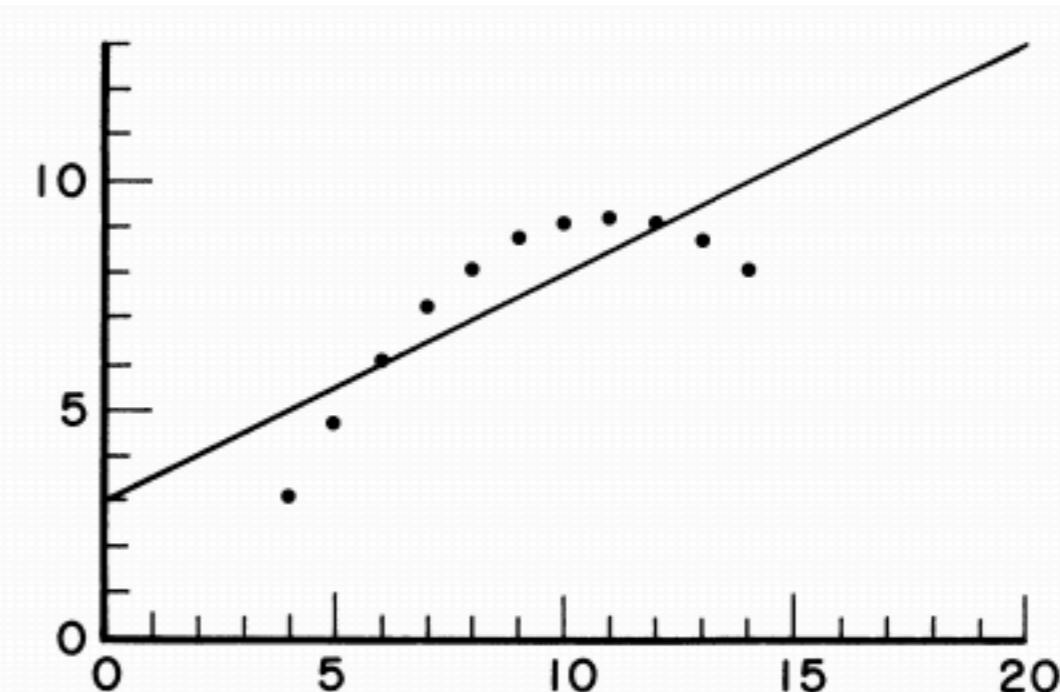
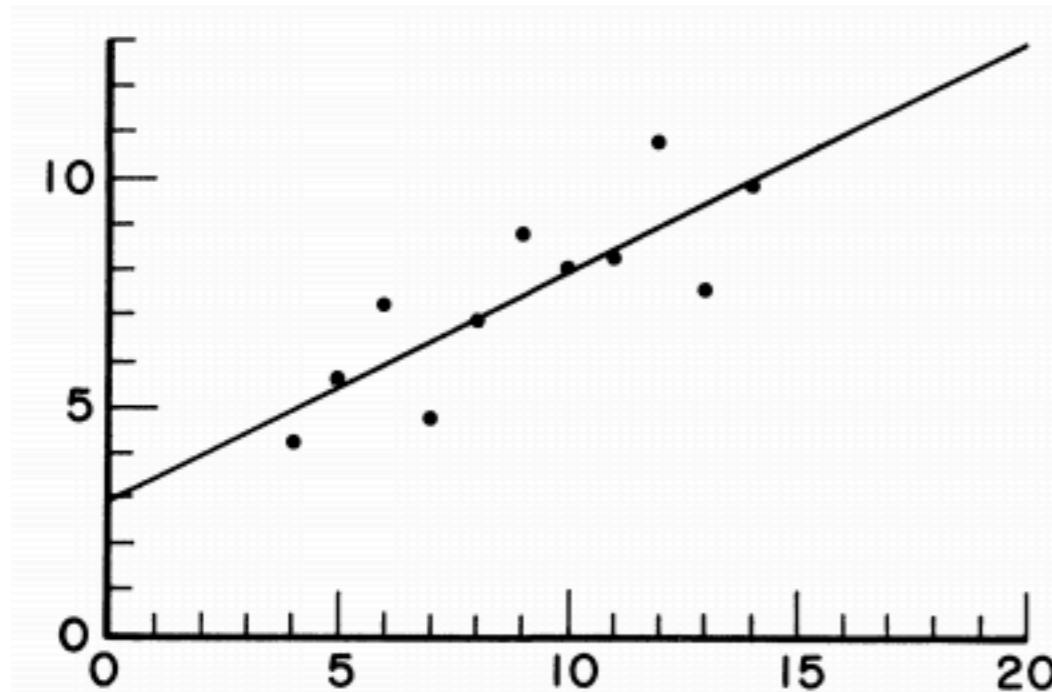
r = 0.816

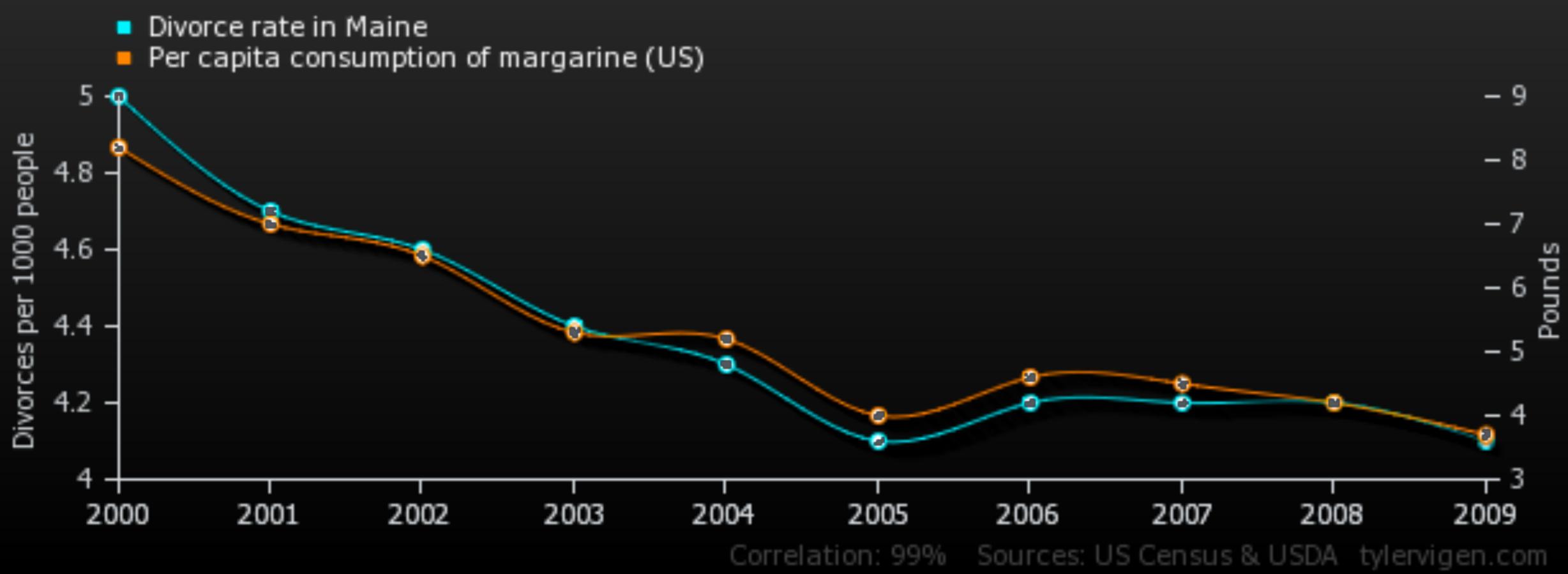
y=3.0+0.5x, r²=0.67

Anscombe (1973)

Data formatting care of Wikipedia



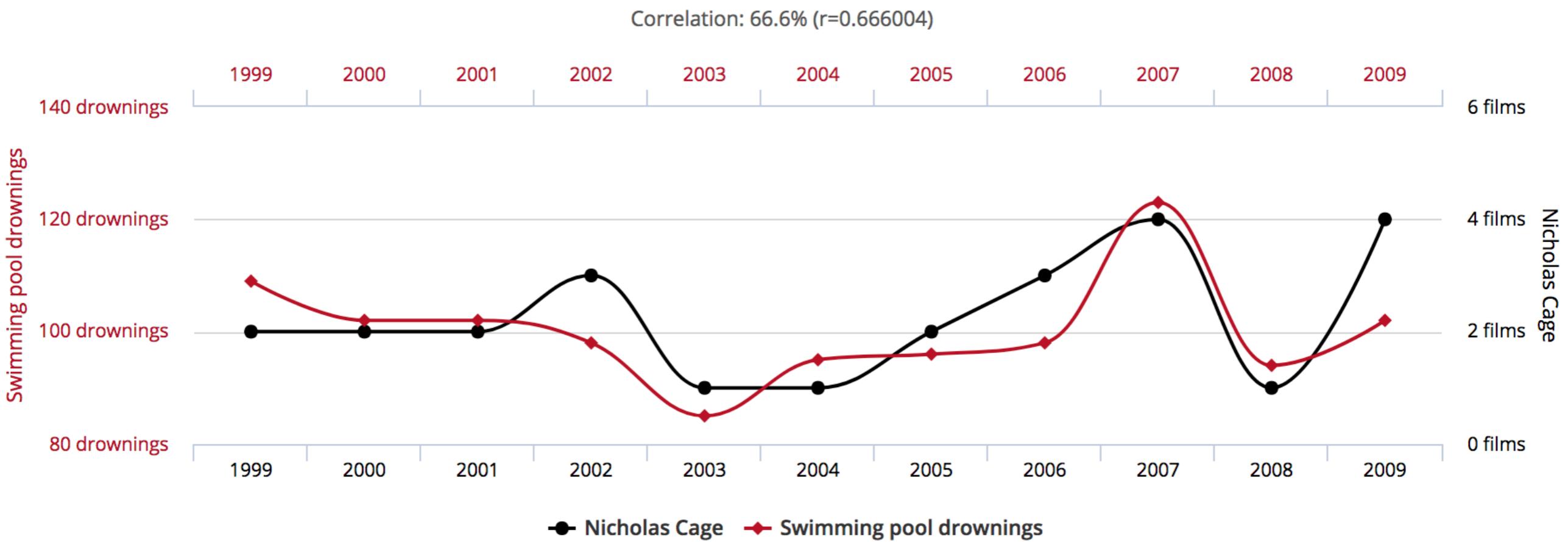




(Teal and orange lines chart divorce rate and veg. butter consumption)



Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in



Data sources: Centers for Disease Control & Prevention and Internet Movie Database

tylervigen.com

(Red and black lines chart drowning deaths and Nic Cage films)



Exposure, the effective laying open of the data to display the **unanticipated**, is to us a major portion of data analysis. Formal statistics has given almost no guidance to exposure; indeed, it is not clear how the **informality and flexibility** appropriate to the **exploratory character of exposure** can be fitted into any of the structures of formal statistics so far proposed.

Nothing - not the careful logic of mathematics, not statistical models and theories, not the awesome arithmetic power of modern computers - nothing can substitute here for the **flexibility** of the **informed human mind**.

Accordingly, both approaches and techniques need to be structured so as to **facilitate human involvement** and **intervention**.

“Data Analytics”

“...the process of **systematically** applying **statistical and/or logical techniques** to **describe** and **illustrate**, **condense** and **recap**, and **evaluate** data.”

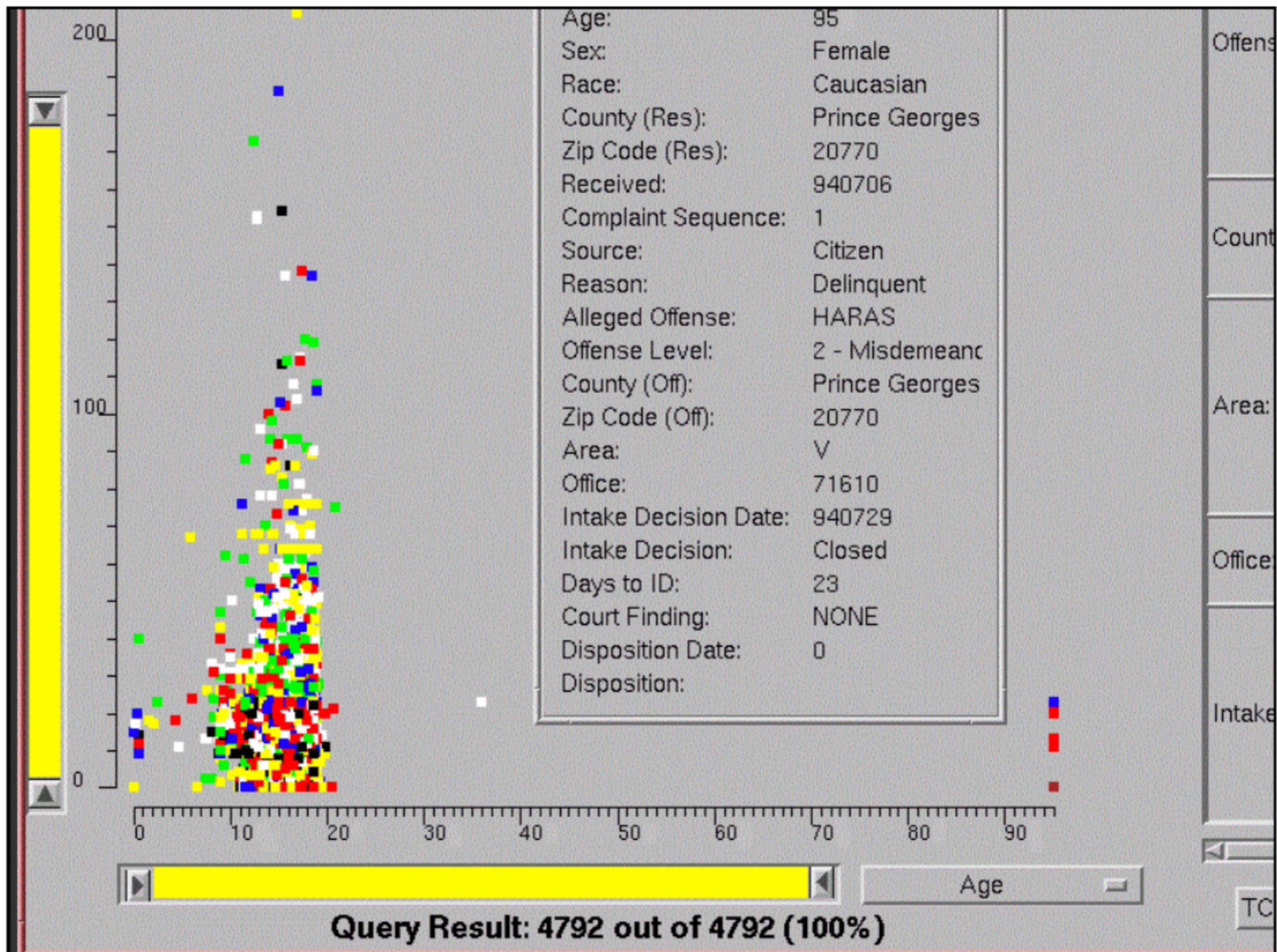
“...provide a way of drawing **inductive inferences** from data and **distinguishing the signal... from the noise** present in the data”

Exploratory Data Analysis

- **Diagnostics** for errors or missing data
- Checking for **common statistical issues** in both the data and your own judgment
- **Pre-visualizing** and **re-representing**
- Transforming, filtering, manipulating, etc

(compare to “confirmatory” analysis)





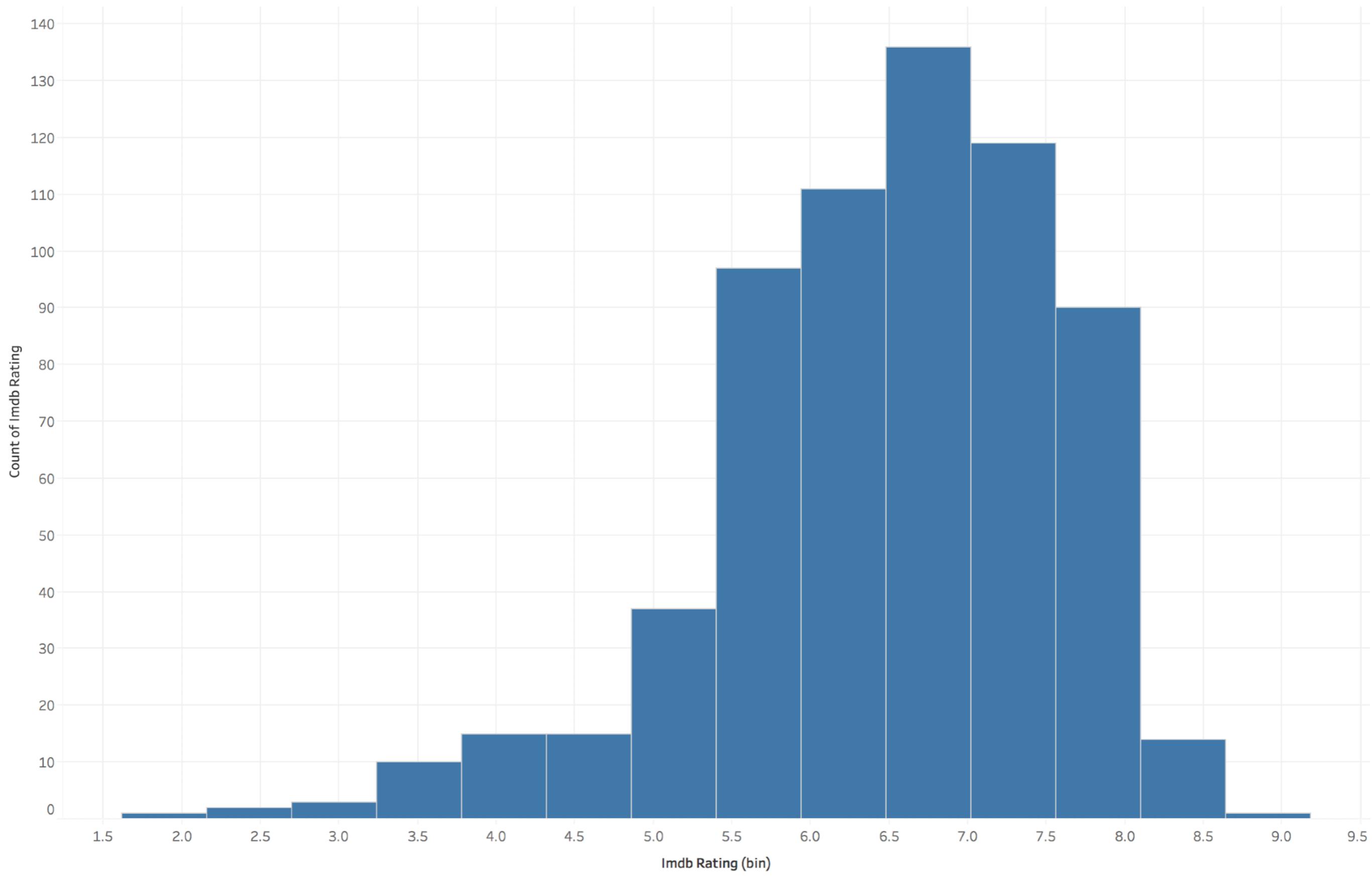
(A starfield visualization shows crime data)



	title	title_type	genre	runtime	mpaa_rating	studio	thtr_rel_year	thtr_rel_month	thtr_rel_day	dvd_rel_year	dvd_rel_month	dvd_rel_day	imdb_rating
1	Filly Brown	Feature Film	Drama	80	R	Indomina Media	2013	4	19	2013	7	30	5.5
2	The Dish	Feature Film	Drama	101	PG-13	Warner Bros.	2001	3	14	2001	8	28	7.3
3	Waiting for Guffman	Feature Film	Comedy	84	R	Sony Pictures	1996	8	21	2001	8	21	7.6
4	The Age of Irrelevance	Feature Film	Drama	139	PG	Columbia Pictures	1993	10	1	2001	11	6	7.2
5	Malevolence	Feature Film	Horror	90	R	Anchor Bay Entertainment	2004	9	10	2005	4	19	5.1
6	Old Partner	Documentary	Documentary	78	Unrated	Shcalo Media	2009	1	15	2010	4	20	7.8
7	Lady Jane	Feature Film	Drama	142	PG-13	Paramount Home Entertainment	1986	1	1	2003	2	18	7.2
8	Mad Dog Tim	Feature Film	Drama	93	R	MGM/United Artists	1996	11	8	2004	3	2	5.5
9	Beauty Is Evil	Documentary	Documentary	88	Unrated	Independent	2012	9	7	2013	1	21	7.5
10	The Snowtov	Feature Film	Drama	119	Unrated	IFC Films	2012	3	2	2012	8	14	6.6
11	Superman II	Feature Film	Action & Adventure	127	PG	Warner Bros.	1981	6	19	2001	5	1	6.8
12	Leap of Faith	Feature Film	Drama	108	PG-13	Paramount Home Entertainment	1992	12	18	2003	9	23	6
13	The Royal Tenenbaums	Feature Film	Comedy	110	R	Buena Vista International	2002	1	4	2002	7	9	7.6
14	School for Scoundrels	Feature Film	Comedy	100	PG-13	MGM	2006	9	23	2007	2	13	5.9
15	Rhinestone	Feature Film	Comedy	111	PG	20th Century Fox	1984	6	20	2005	3	22	3.6
16	Burn After Reading	Feature Film	Drama	96	R	Focus Features	2008	8	27	2008	12	21	7
17	The Doors	Feature Film	Drama	140	R	Sony Pictures	1991	3	1	1997	8	19	7.2
18	The Wood	Feature Film	Drama	106	R	Paramount Pictures	1999	7	16	2000	1	18	6.9
19	Jason X	Feature Film	Horror	91	R	New Line Cinema	2002	4	26	2004	6	1	4.4
20	Dragon Wars	Feature Film	Drama	90	PG-13	Sony Pictures	2007	9	13	2008	1	8	3.6
21	Fallen	Feature Film	Drama	124	R	Warner Home Entertainment	1997	6	1	1998	6	16	7
22	The Gleaners	Documentary	Documentary	82	Unrated	Zeitgeist Film	2001	4	6	2002	7	23	7.6
23	Tamara Drewe	Feature Film	Drama	107	R	Sony Pictures	2010	10	8	2011	2	8	6.2
24	Not Without My Daughter	Feature Film	Drama	116	PG-13	MGM Home Entertainment	1991	1	11	2001	10	2	6.2
25	The Yes Men	Documentary	Documentary	87	Unrated	Cinetic Media	2009	1	18	2010	4	1	7.6
26	Oh, God! You've Seen One	Feature Film	Comedy	97	PG	WARNER BROS.	1984	11	9	2005	2	15	5.3
27	Meet Joe Black	Feature Film	Drama	178	PG-13	Universal Pictures	1998	11	13	1999	4	6	7.1
28	The Bay	Feature Film	Horror	84	R	Roadside Attractions	2012	11	2	2013	3	5	5.6
29	Scandal	Feature Film	Art House & Indie	115	R	HBO Video	1989	5	28	2000	9	5	6.4
30	Imagine: John Lennon	Documentary	Musical & Performance	100	R	WARNER BROS.	1988	10	7	2005	12	6	7.9
31	Locke	Feature Film	Mystery & Suspense	85	R	A24 Films	2014	4	25	2014	8	12	7.1
32	The SpongeBob SquarePants Movie	Feature Film	Animation	87	PG	Paramount Pictures	2004	11	19	2005	3	1	7
33	The Muppets	Feature Film	Science Fiction	85	G	Walt Disney	1992	1	1	2002	10	8	7.7
34	Die Hard 2	Feature Film	Mystery & Suspense	124	R	20th Century Fox	1990	7	2	1999	3	9	7.1
35	The Reckoning	Feature Film	Documentary	112	R	Paramount Classics	2004	3	5	2004	8	3	6.6
36	Alice in Wonderland	Feature Film	Action & Adventure	108	PG	Walt Disney	2010	3	5	2010	6	1	6.5



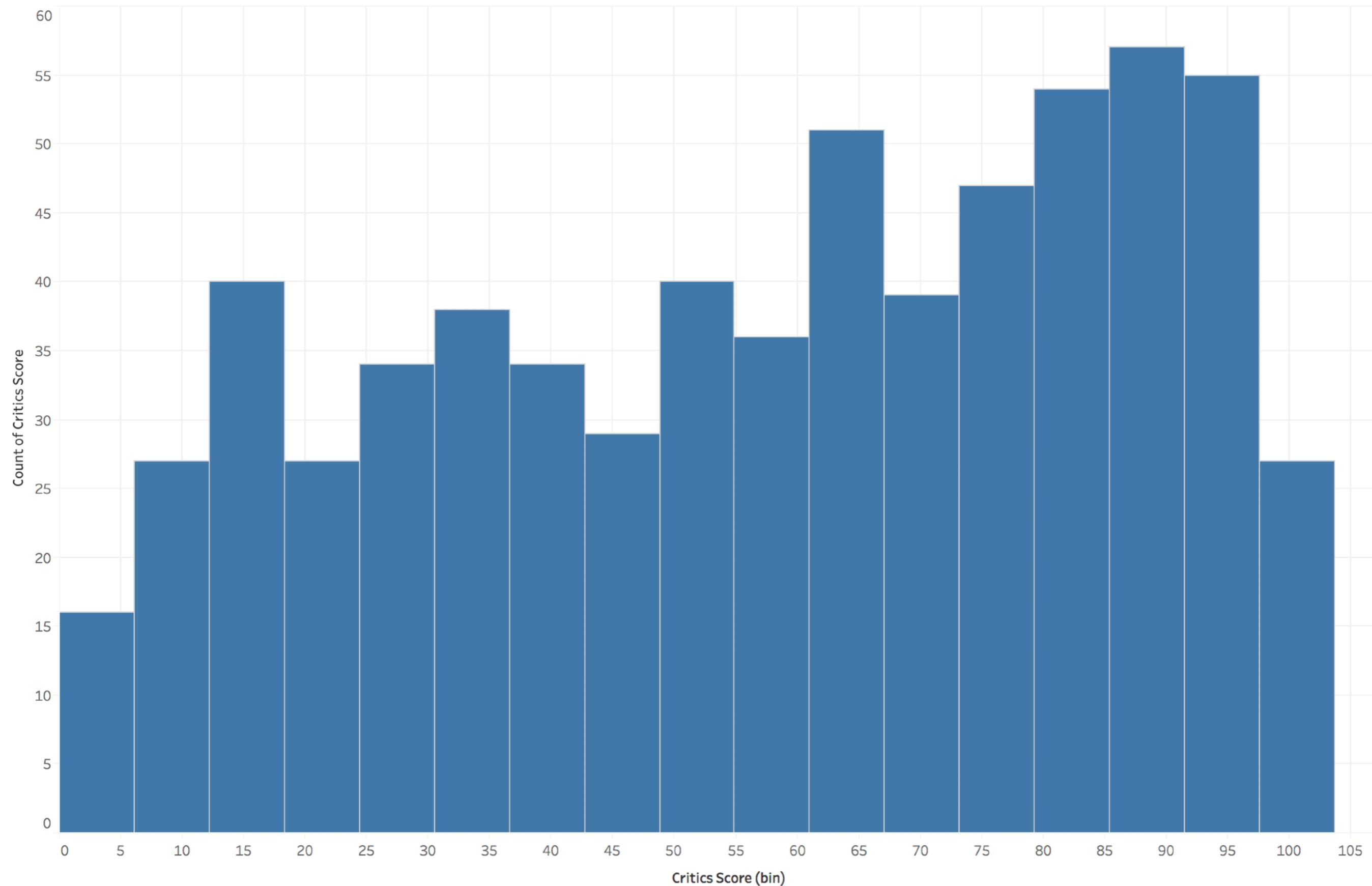
(all done in Tableau)



(Marks: boxes Channels: aligned pos & len)



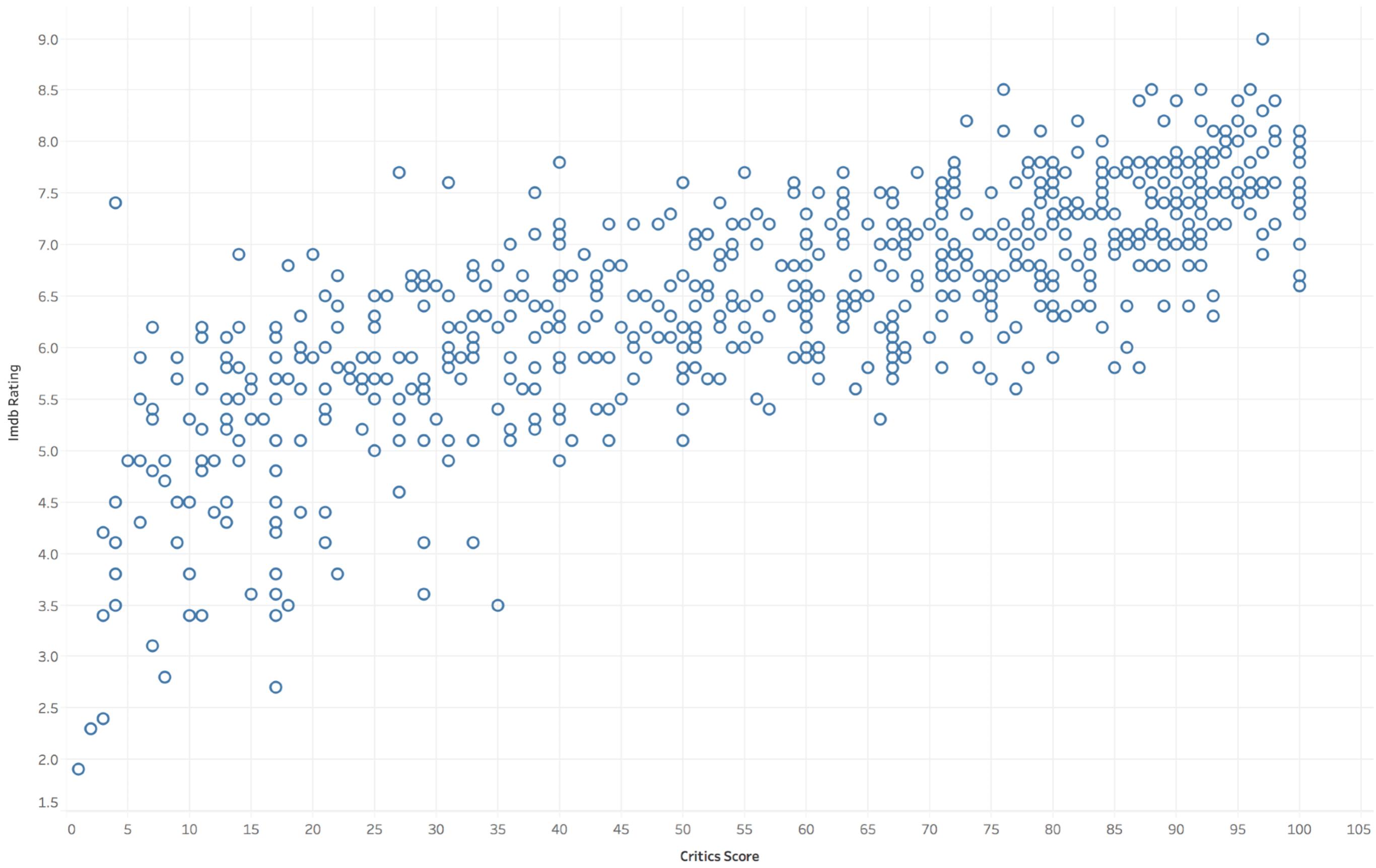
(all done in Tableau)



(Marks: boxes Channels: aligned pos & len)

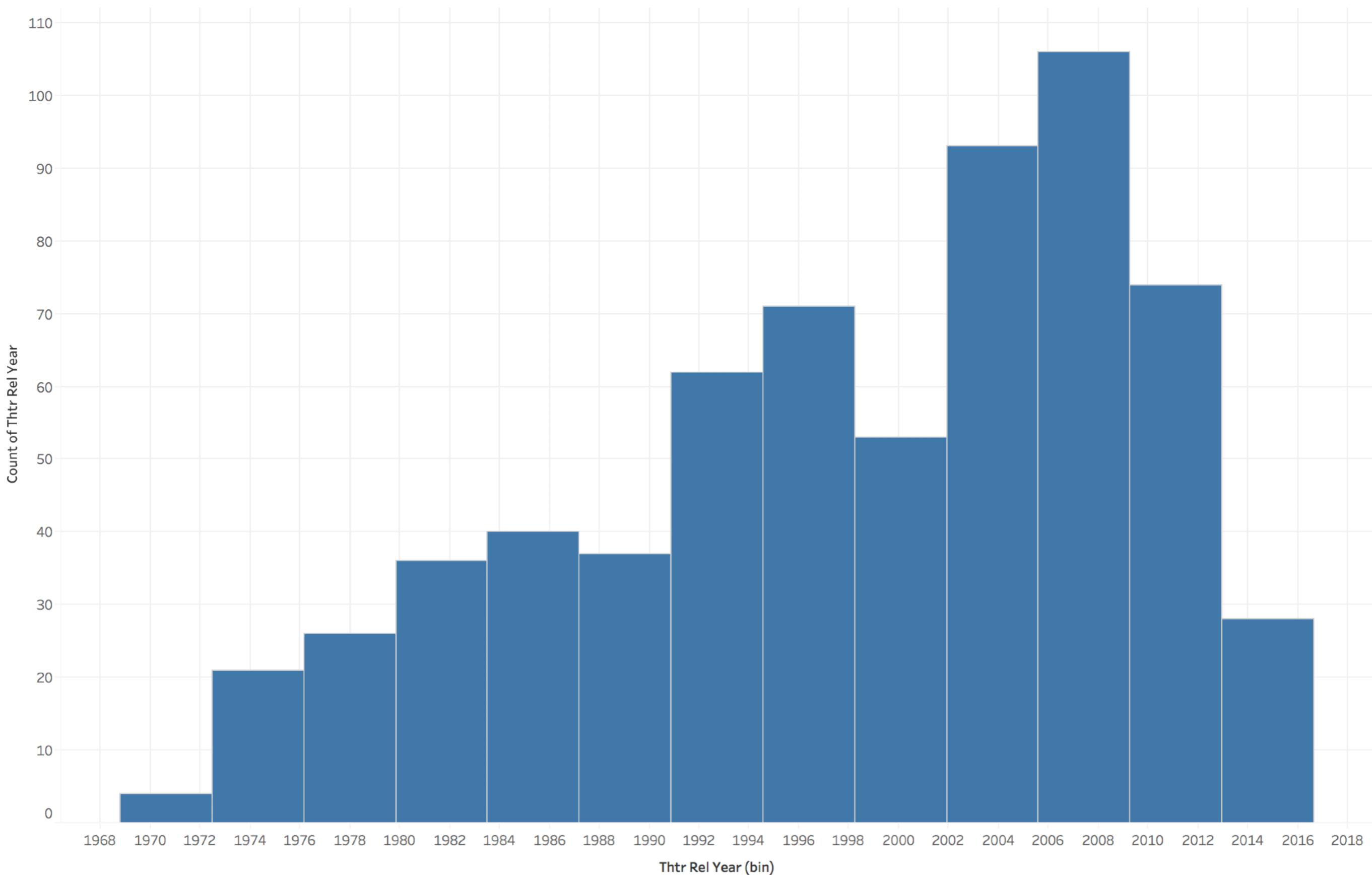


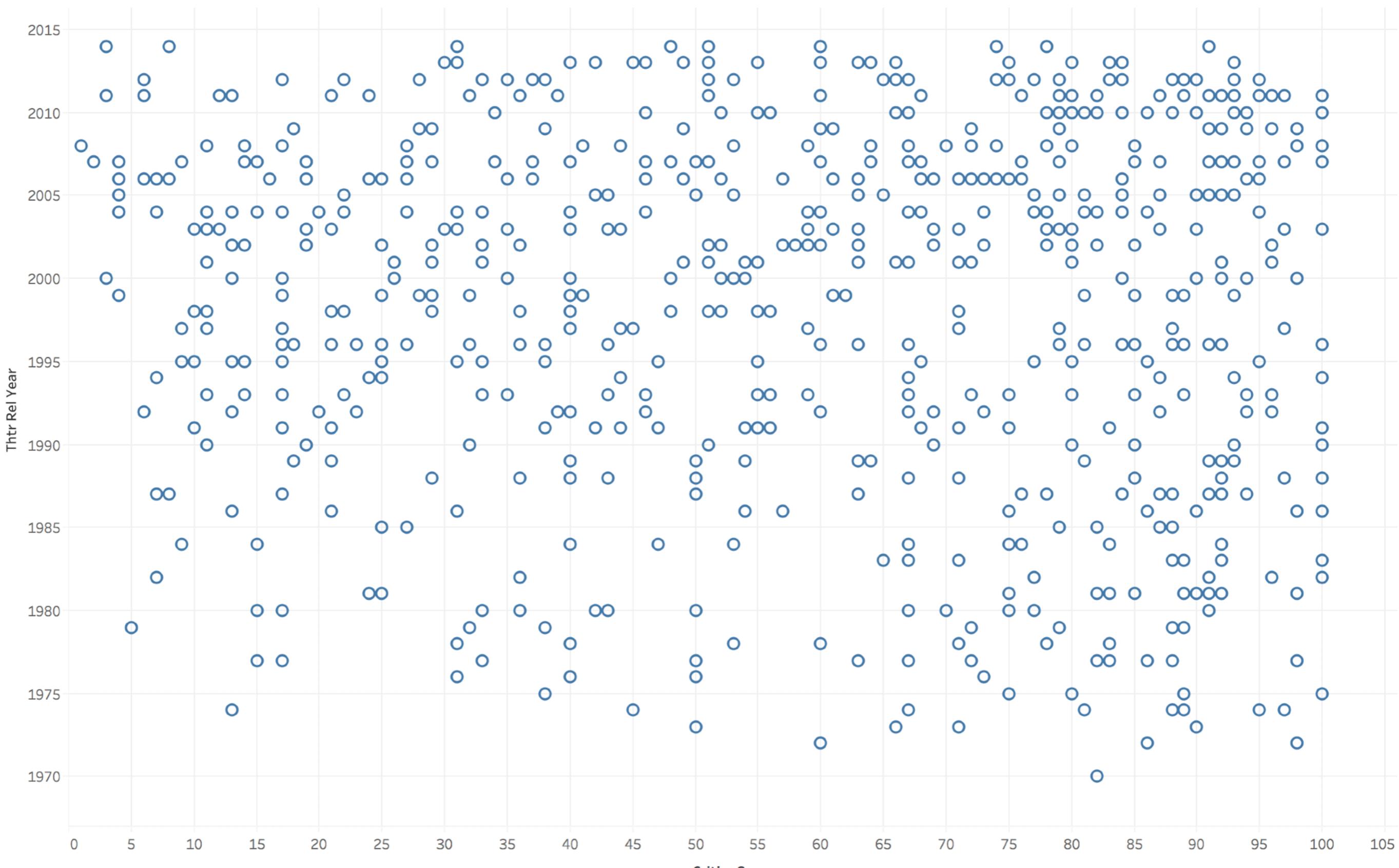
(all done in Tableau)



(Marks: circles Channels: aligned pos x 2)



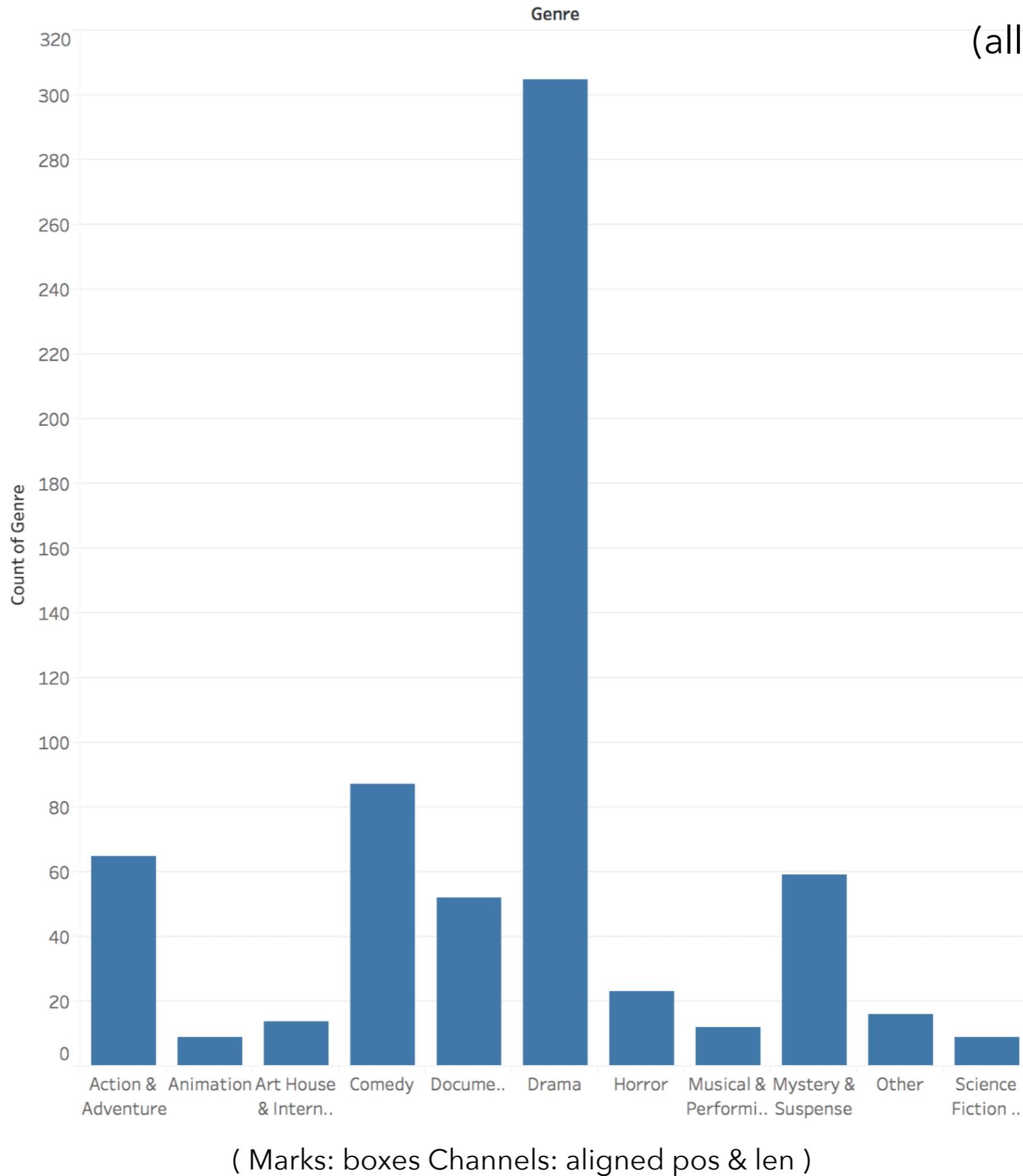




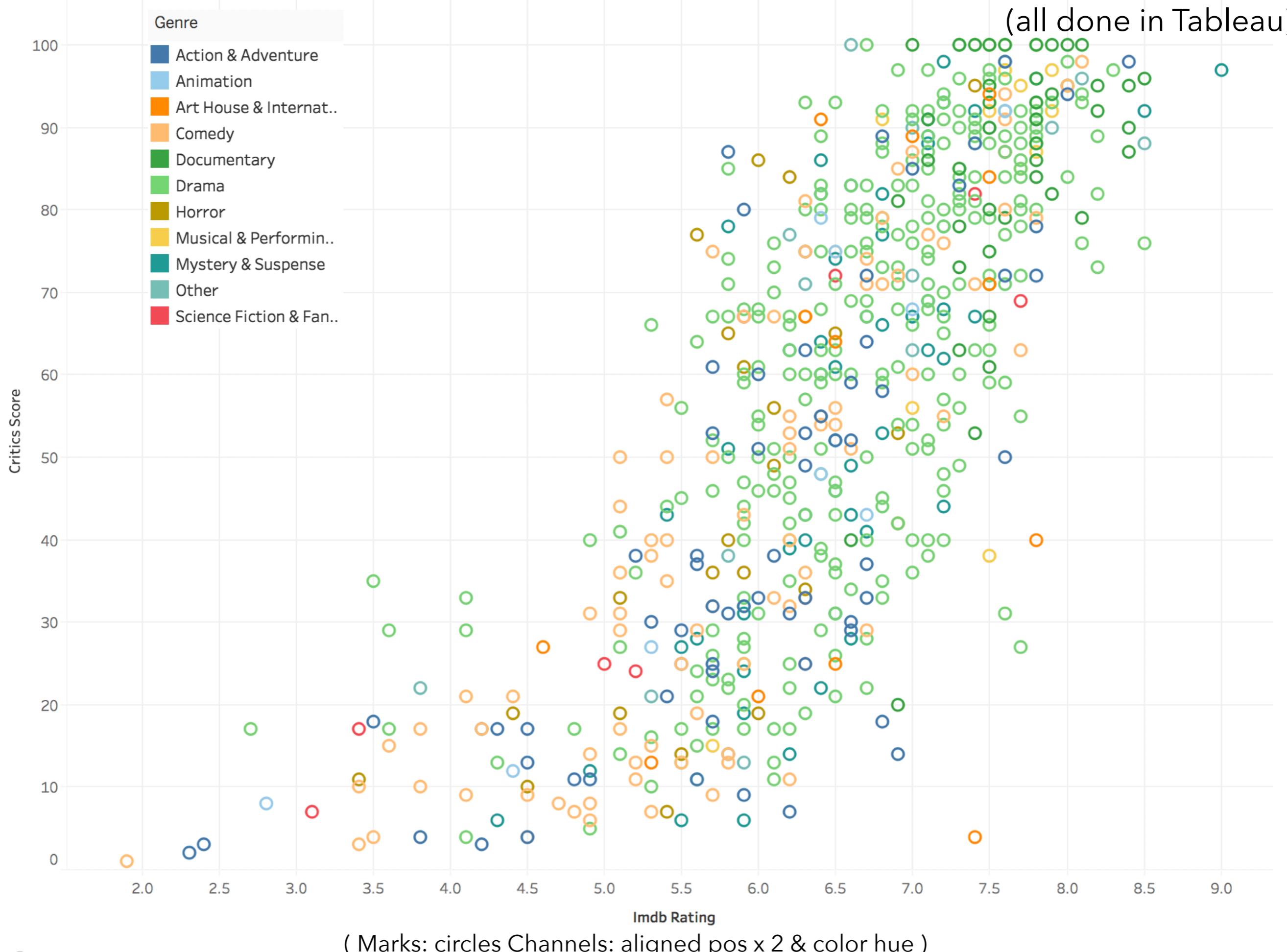
(Marks: circles Channels: aligned pos x 2)



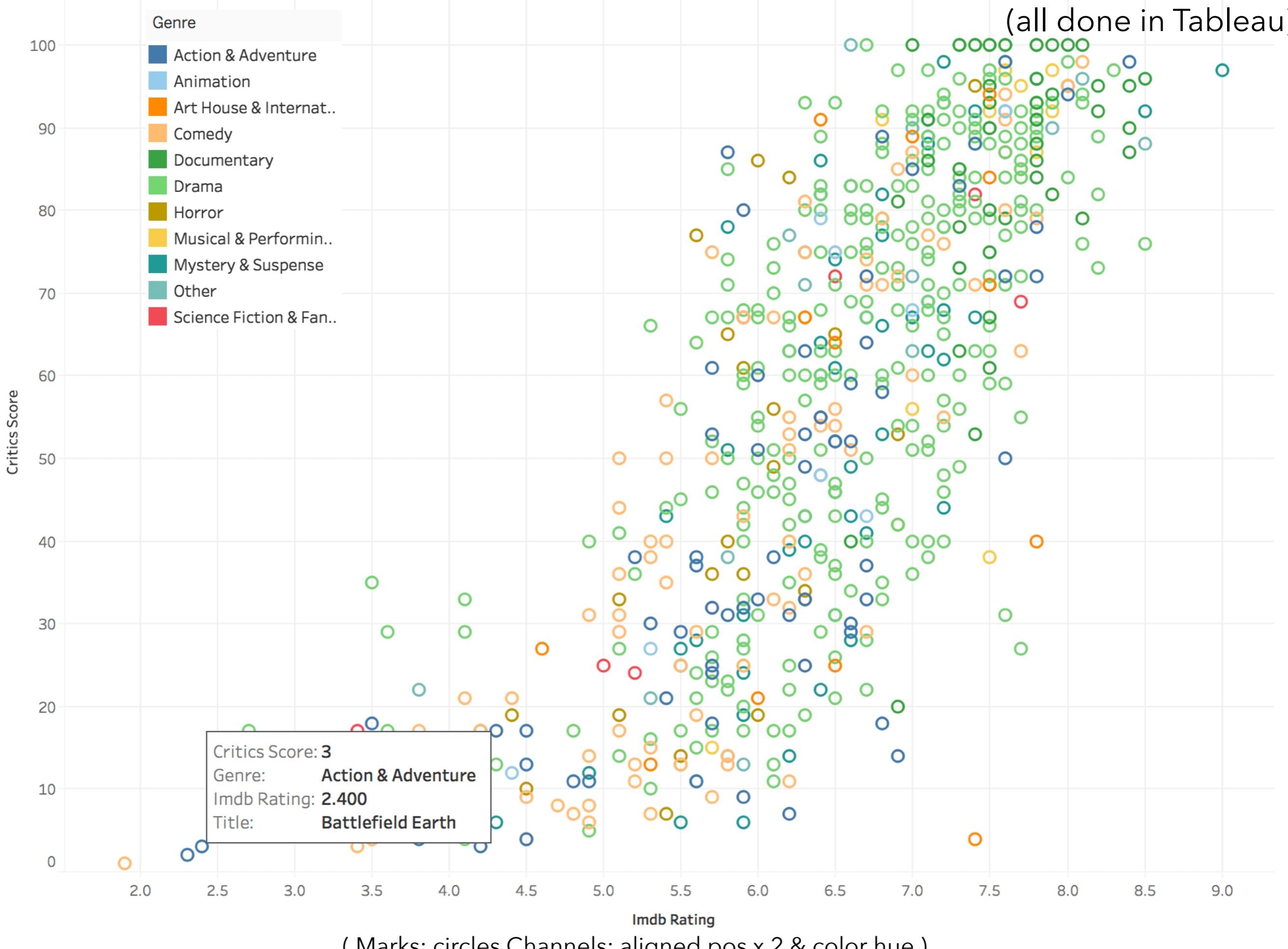
(all done in Tableau)



(all done in Tableau)



(all done in Tableau)



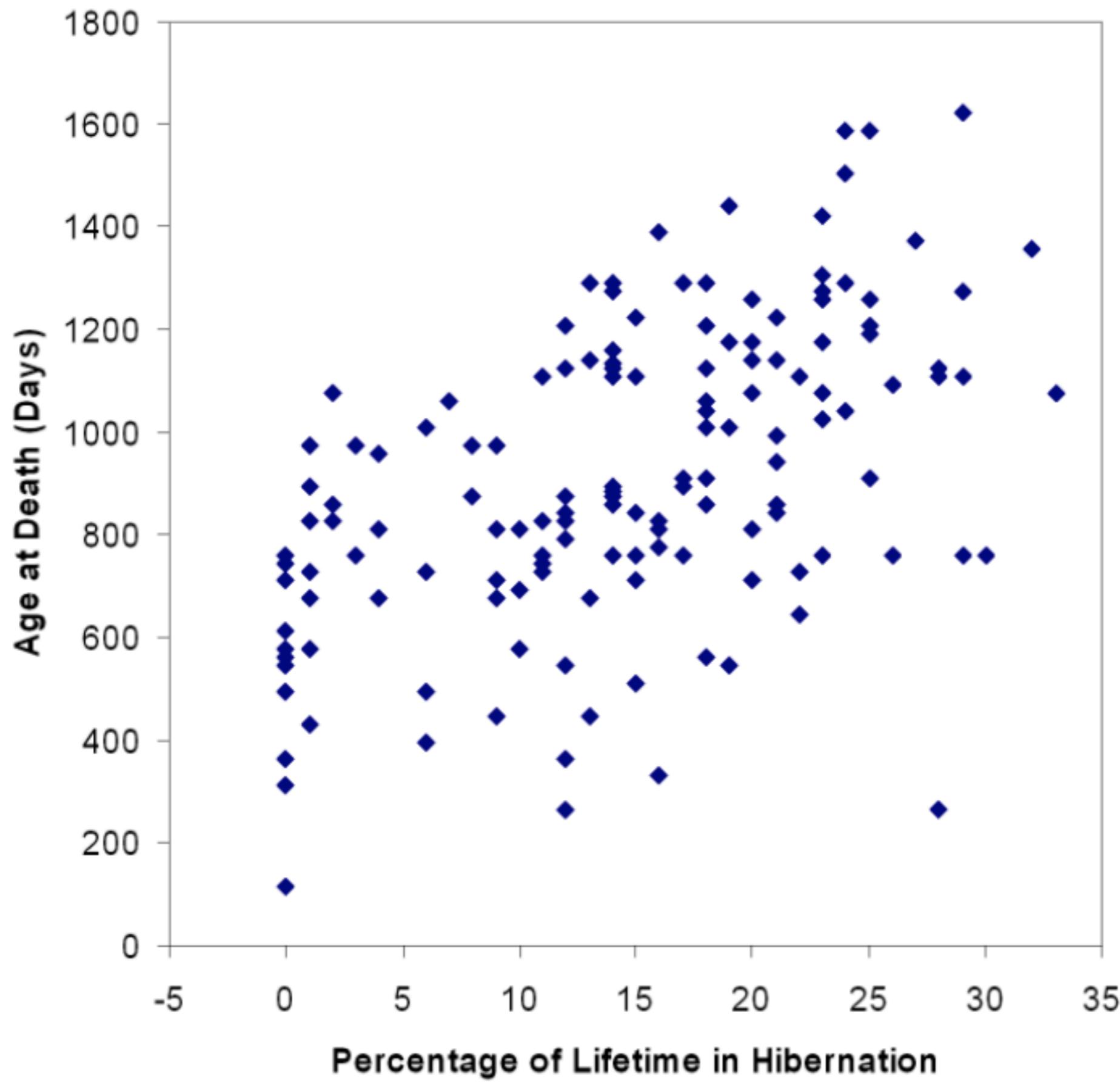
1. Spot check quality
2. Summarize values, then look at relationships
3. Identify trends and targets (but don't fixate early!)

Exploration Sanity Checks

Make sure that your statistical tests
are returning good data
(this is where outliers will get you)

Apply confirmatory analysis
(where appropriate)

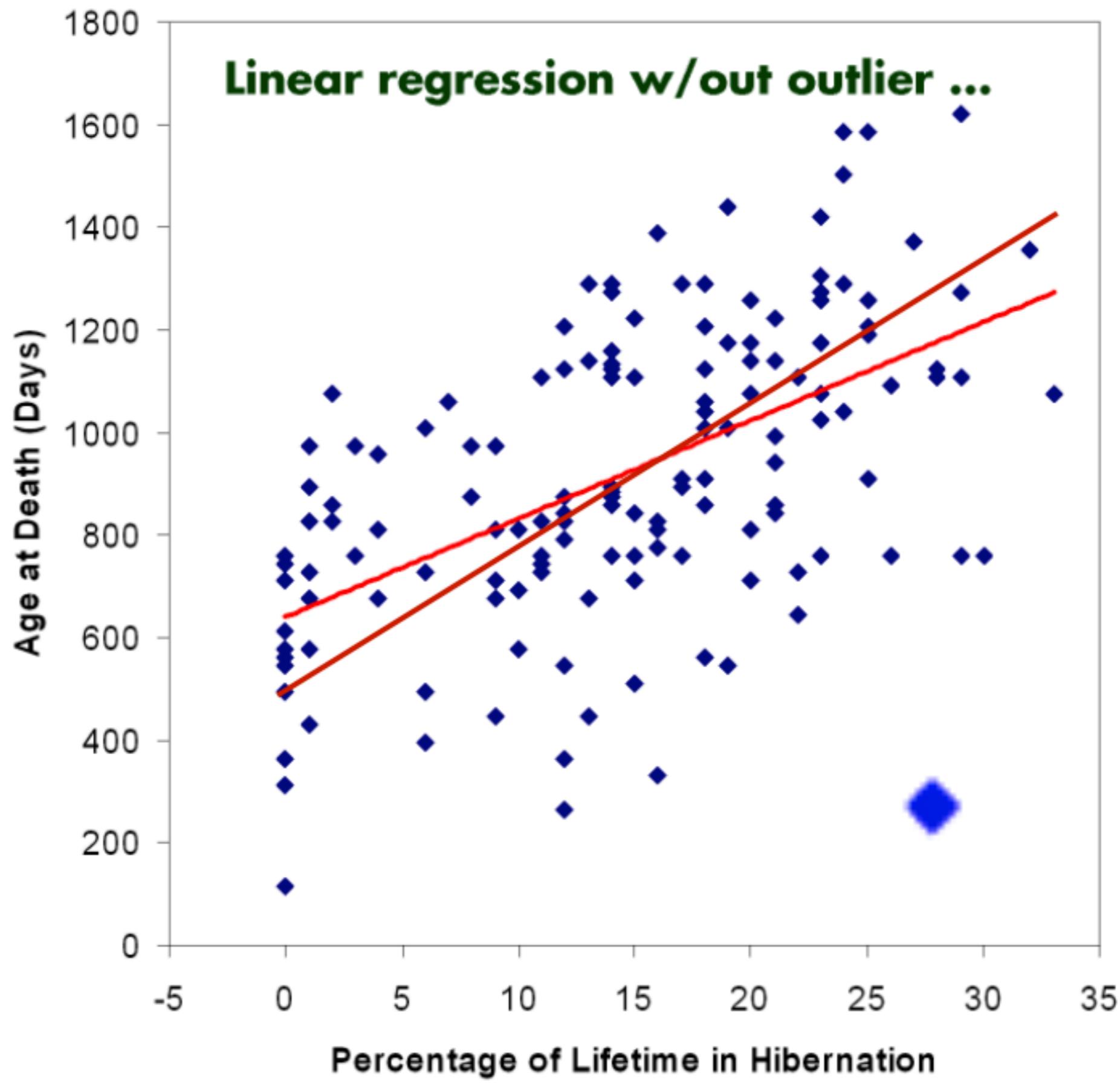




(Marks: diamonds Channels: aligned pos x 2)

Cleveland, 1994





(Marks: diamonds & lines Channels: aligned pos x 2)

Cleveland, 1994



Confirmatory Data Analysis

- “How likely is it that the pattern I’m seeing is just random chance”
- Formal statistical tests
- Graphical inference

