



10

DISTRIBUIÇÃO

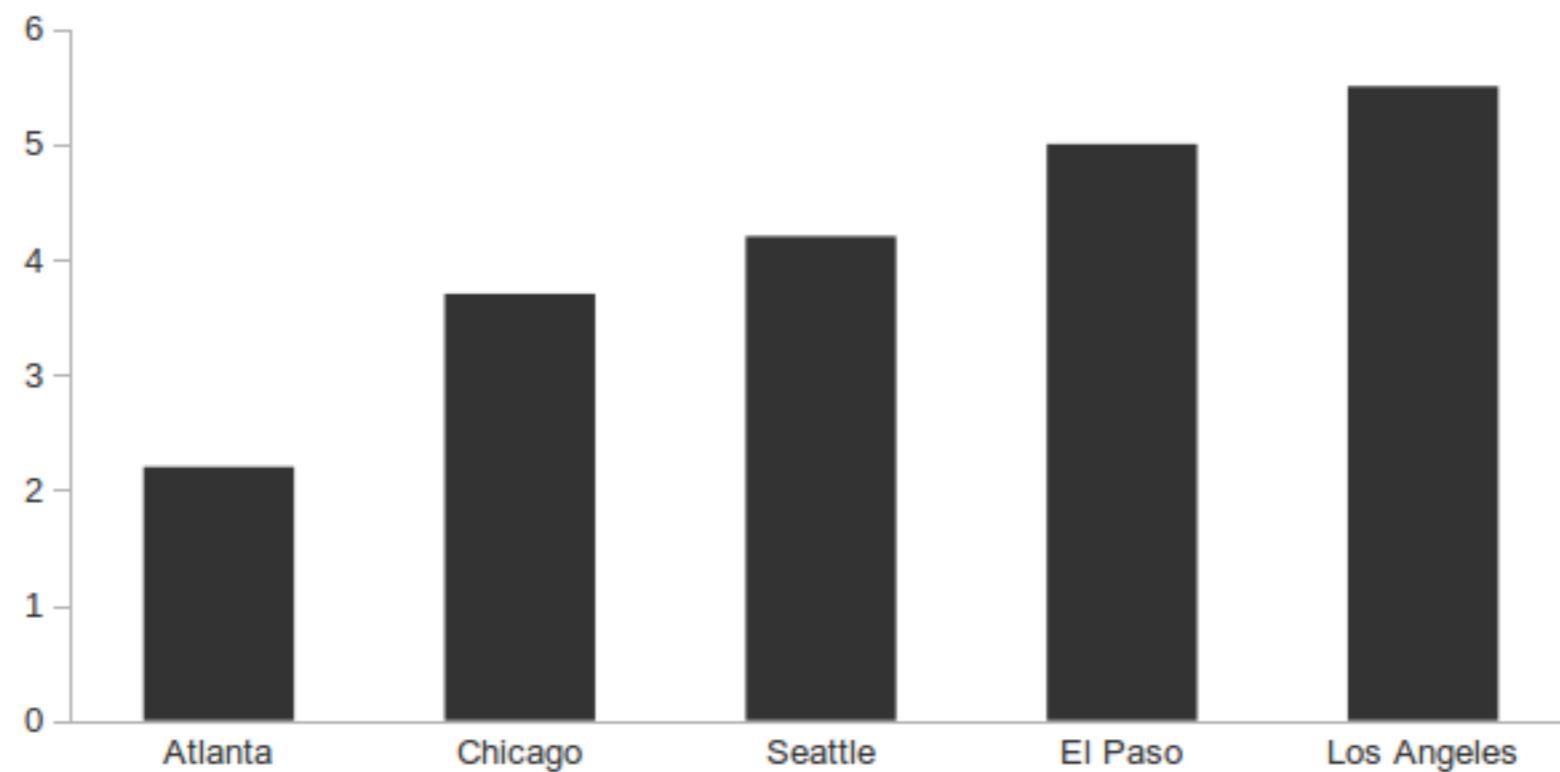
Profa. Raquel C. de Melo Minardi

- Análise de como um conjunto de valores está distribuído em um intervalo

ou ainda

- Contrastar como um ou mais conjunto de valores se distribuem

Average Days to Ship Orders

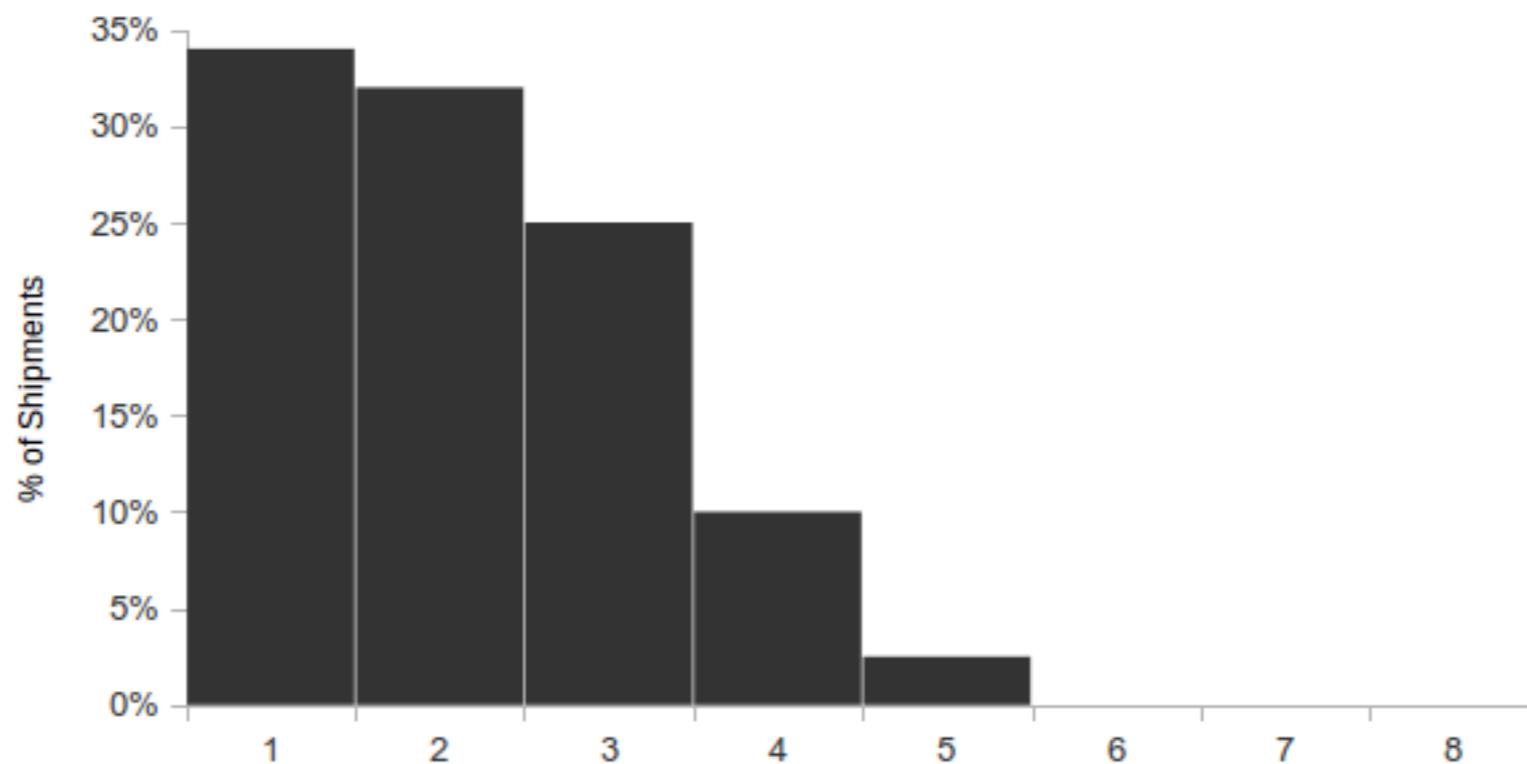


1 1 1 2 2 2 3 3 3 3 4 4 5 5 5 6 7 7 8 8

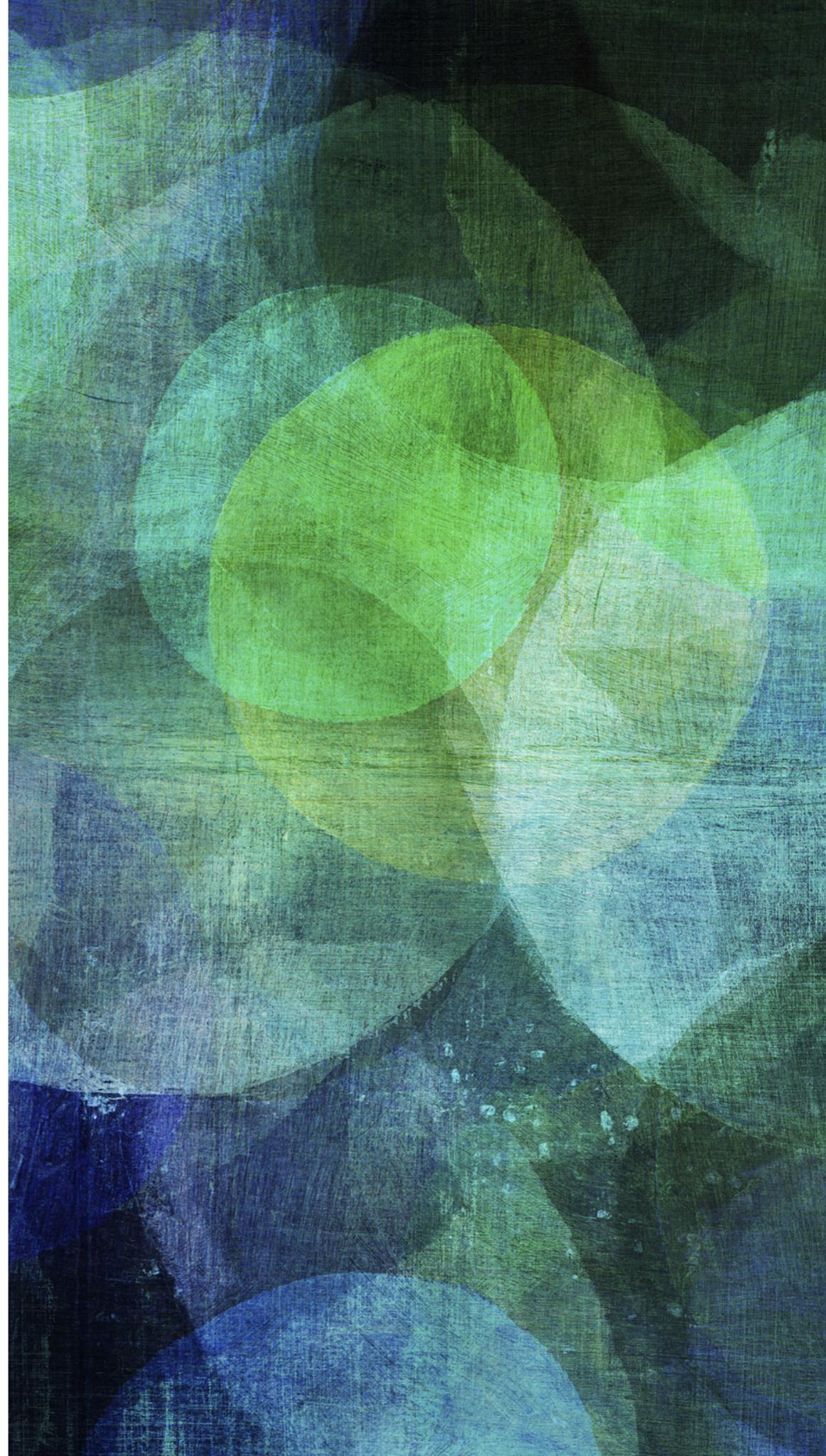
Seattle



Atlanta

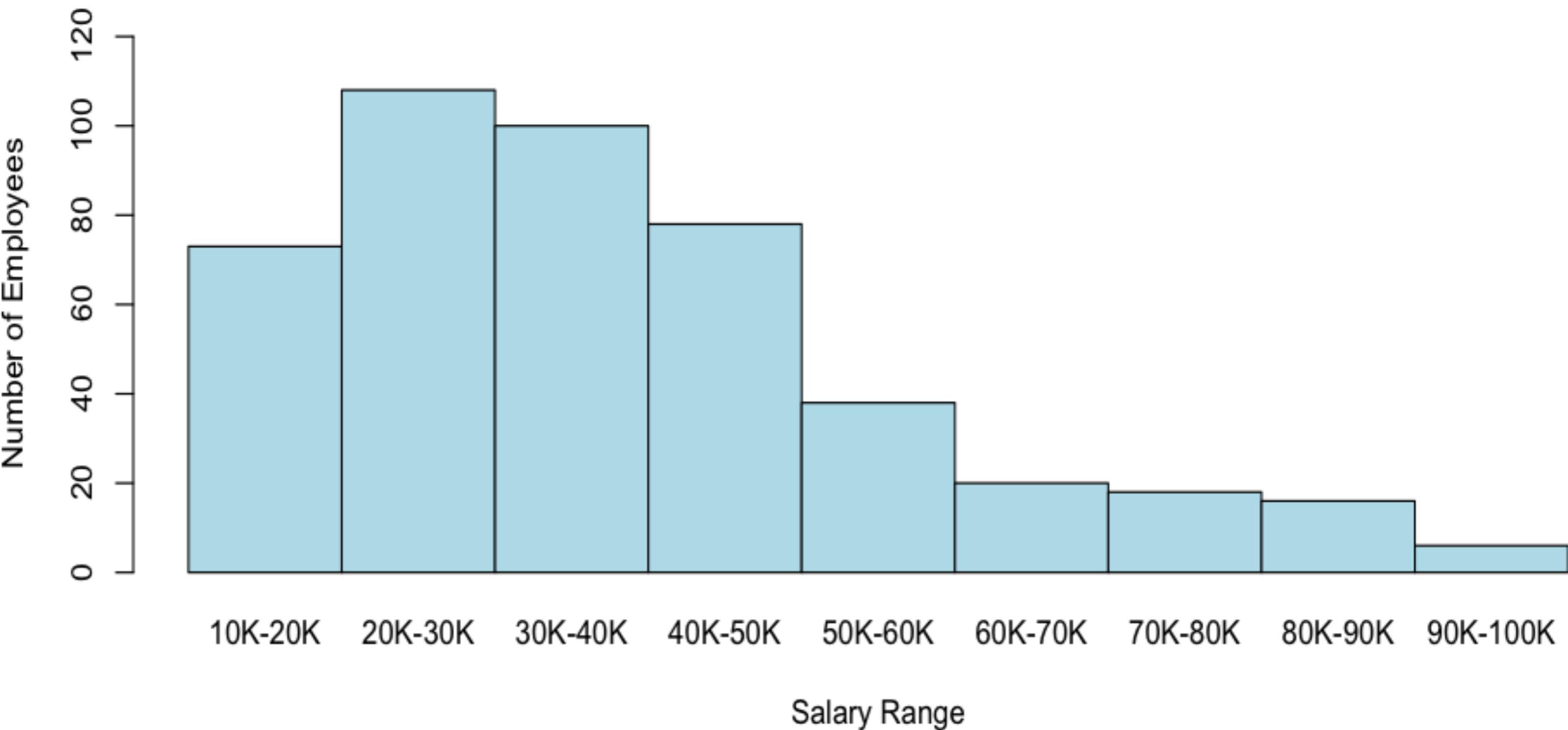


CARACTERÍSTICAS VISUAIS DE DISTRIBUIÇÕES

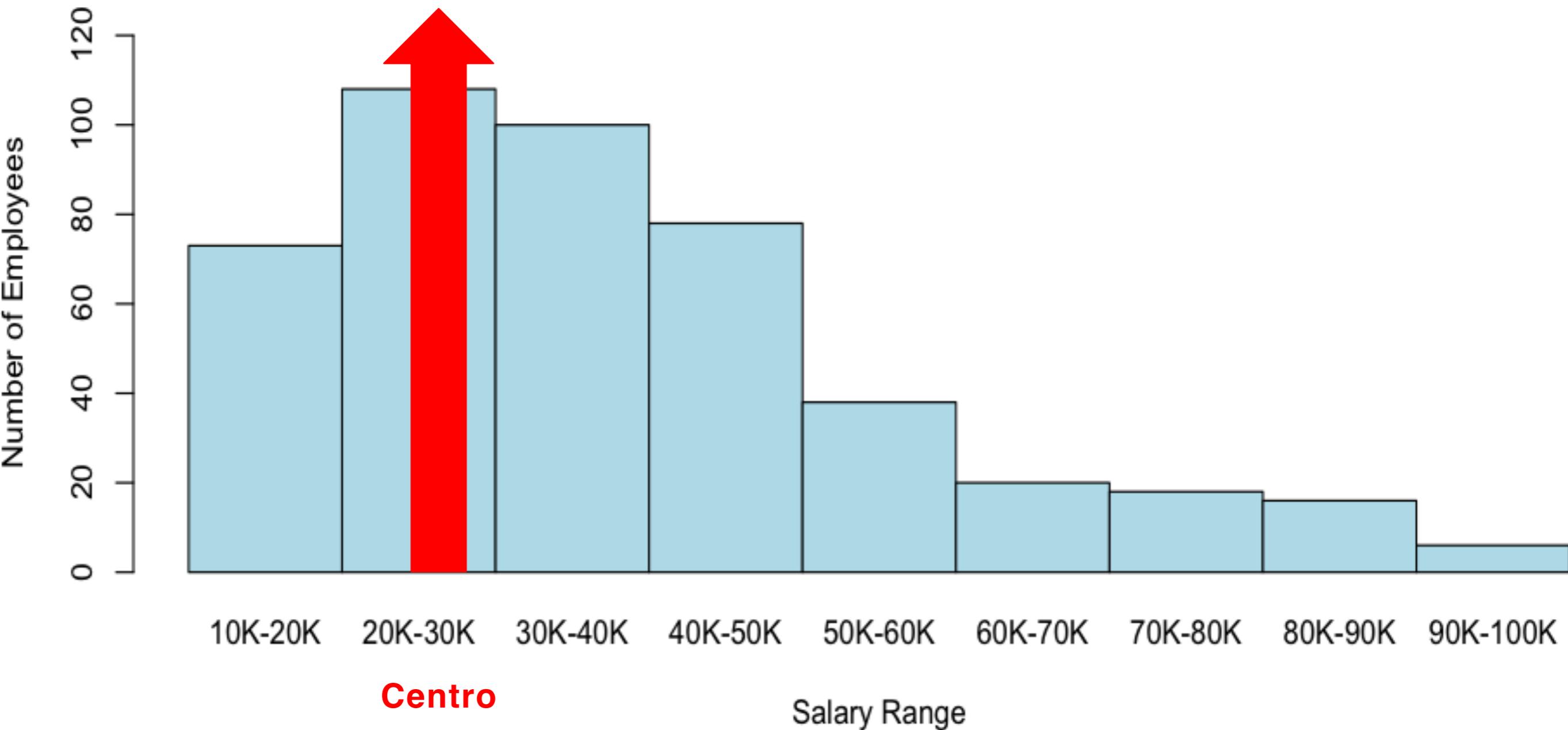


CARACTERÍSTICAS VISUAIS DE DISTRIBUIÇÕES

- Espalhamento
- Centro
- Forma



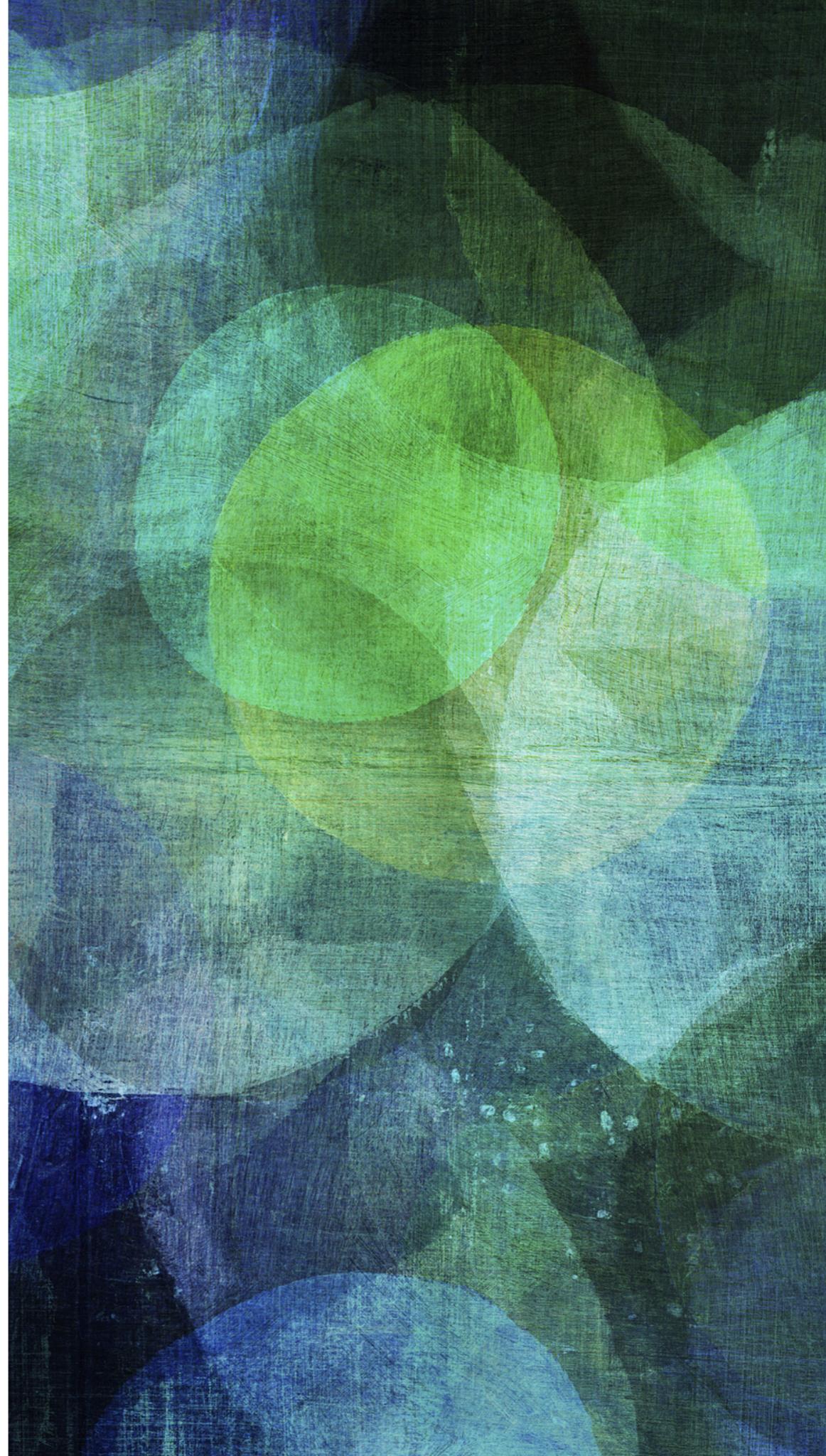
Espalhamento



Forma

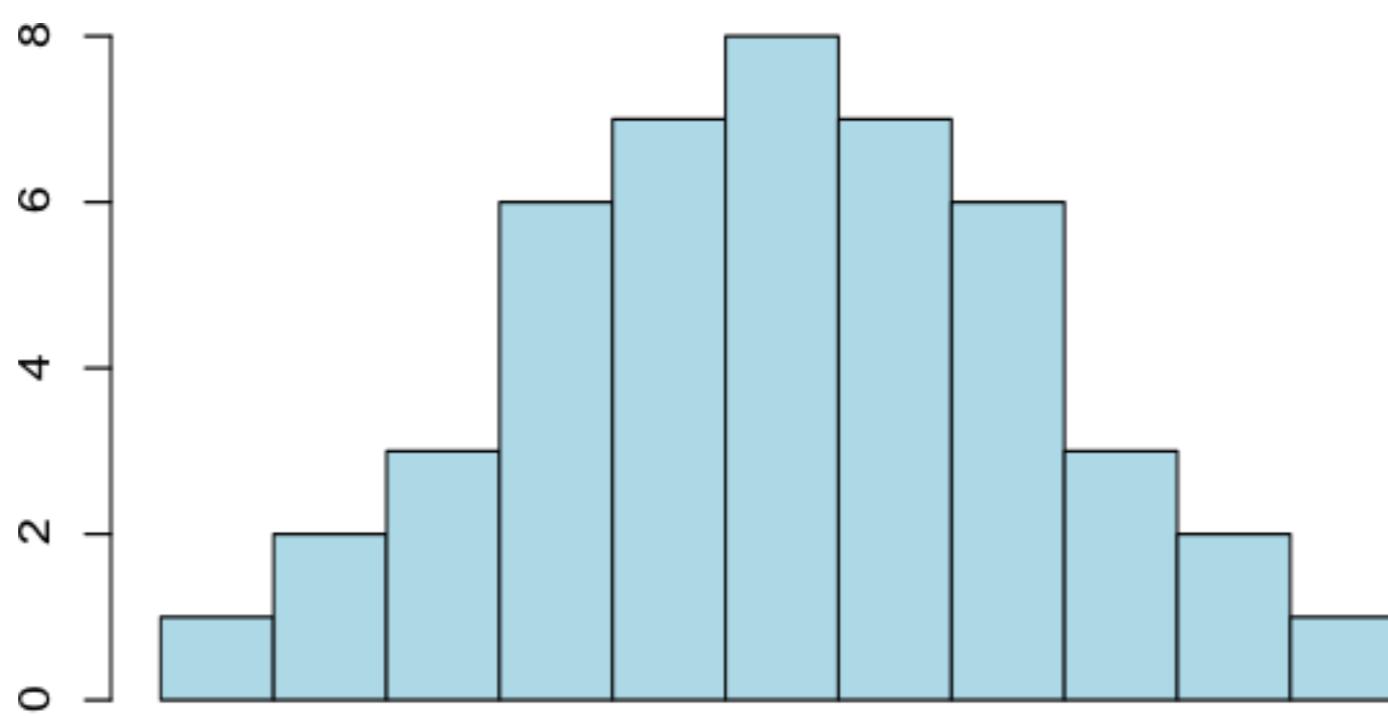


PADRÕES ANALÍTICOS

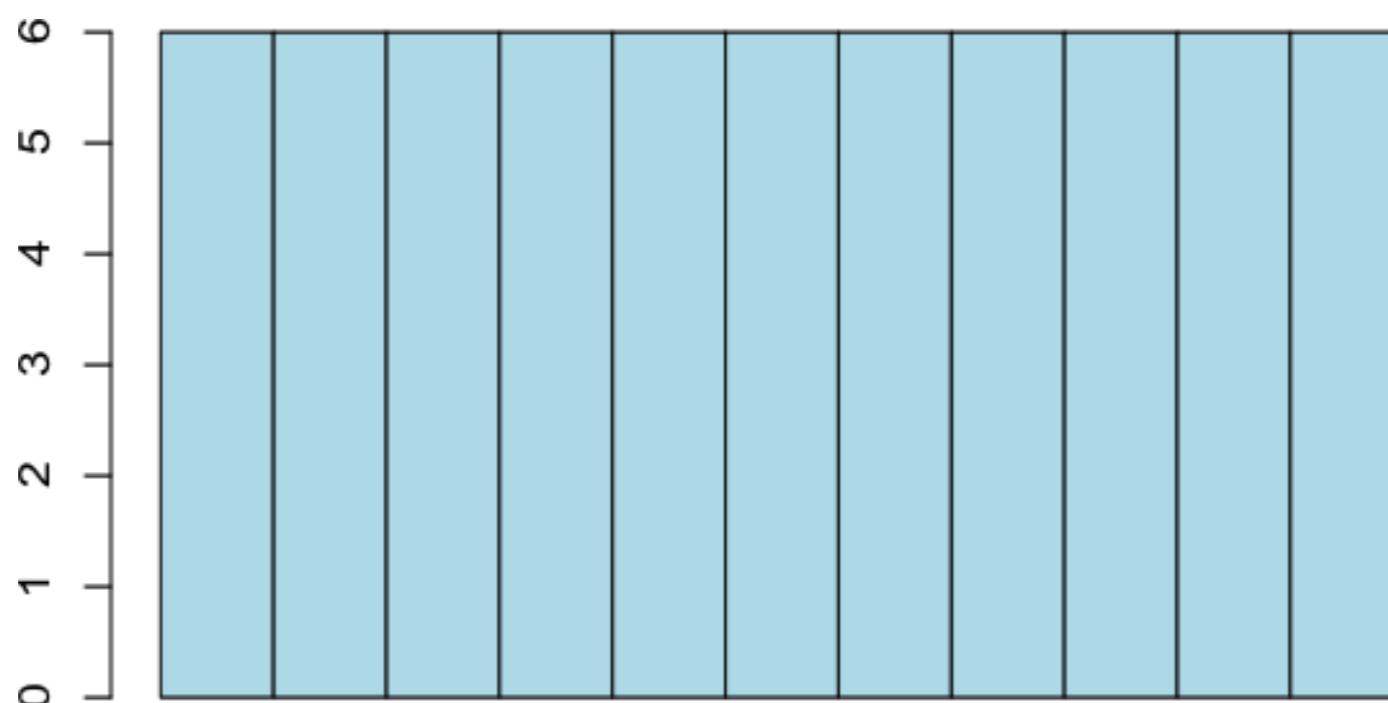


PADRÕES ANALÍTICOS

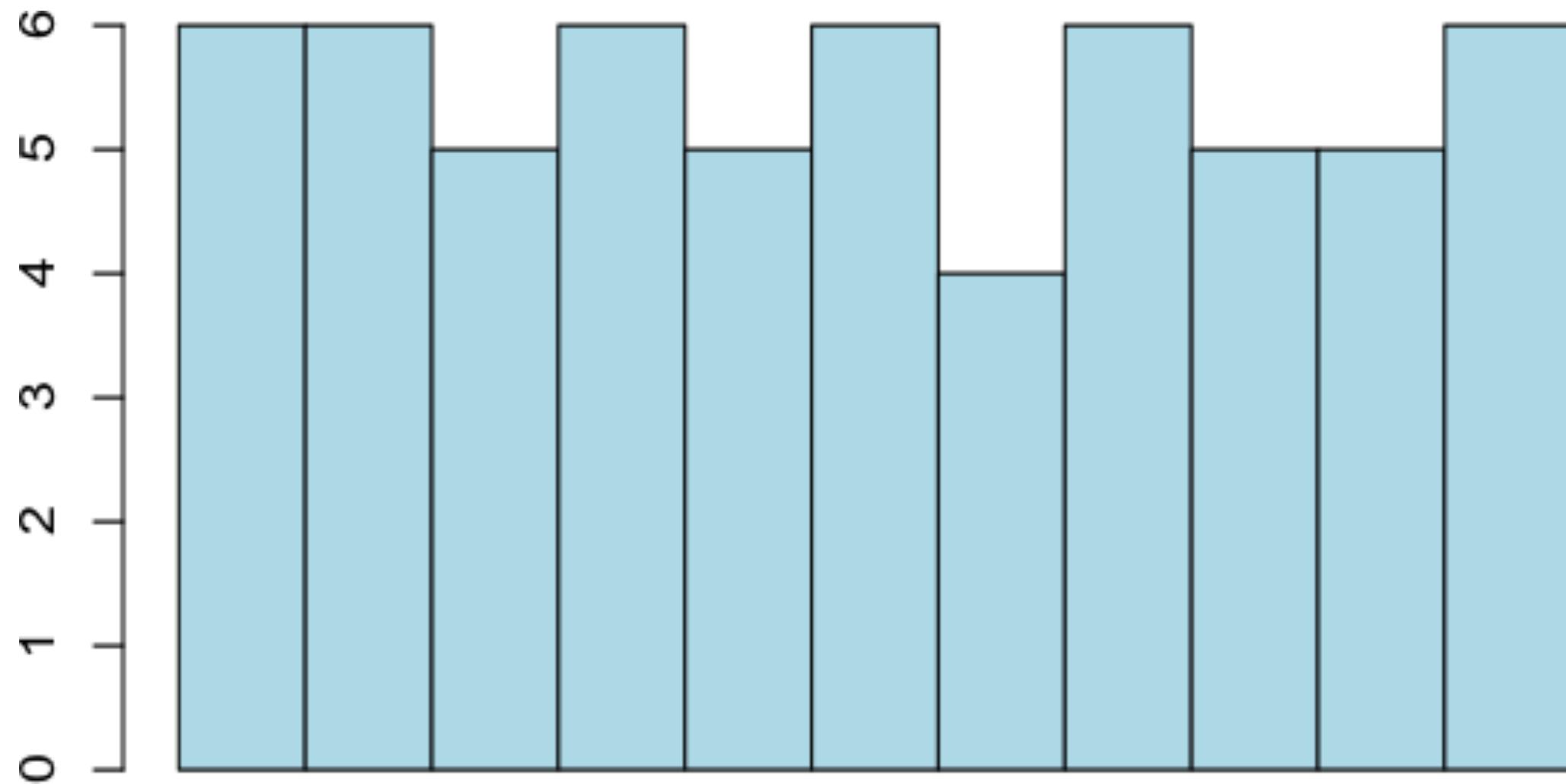
- Curva ou plana?
- Se curva, crescente ou decrescente?
- Se curva crescente, um ou vários picos?
- Se apenas um pico, simétrica ou enviesada?
- Concentrações?
- Lacunas?

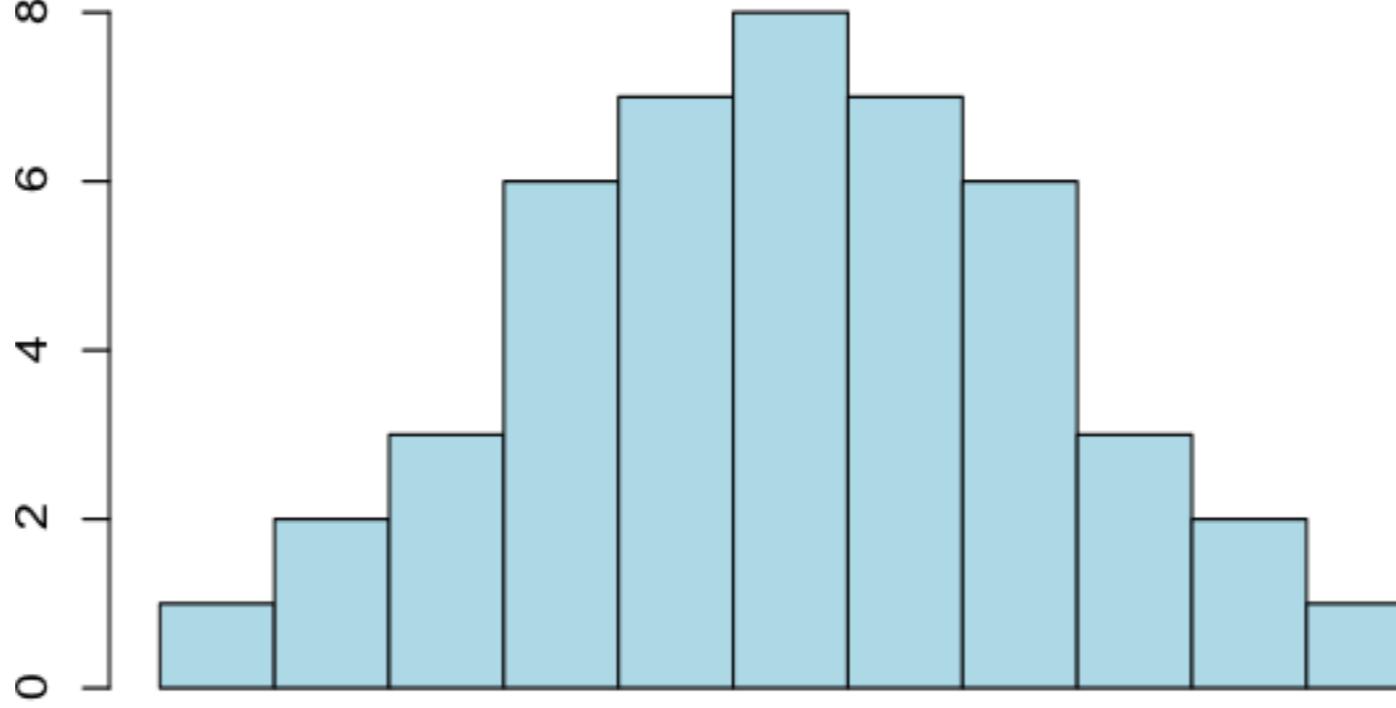


Curved

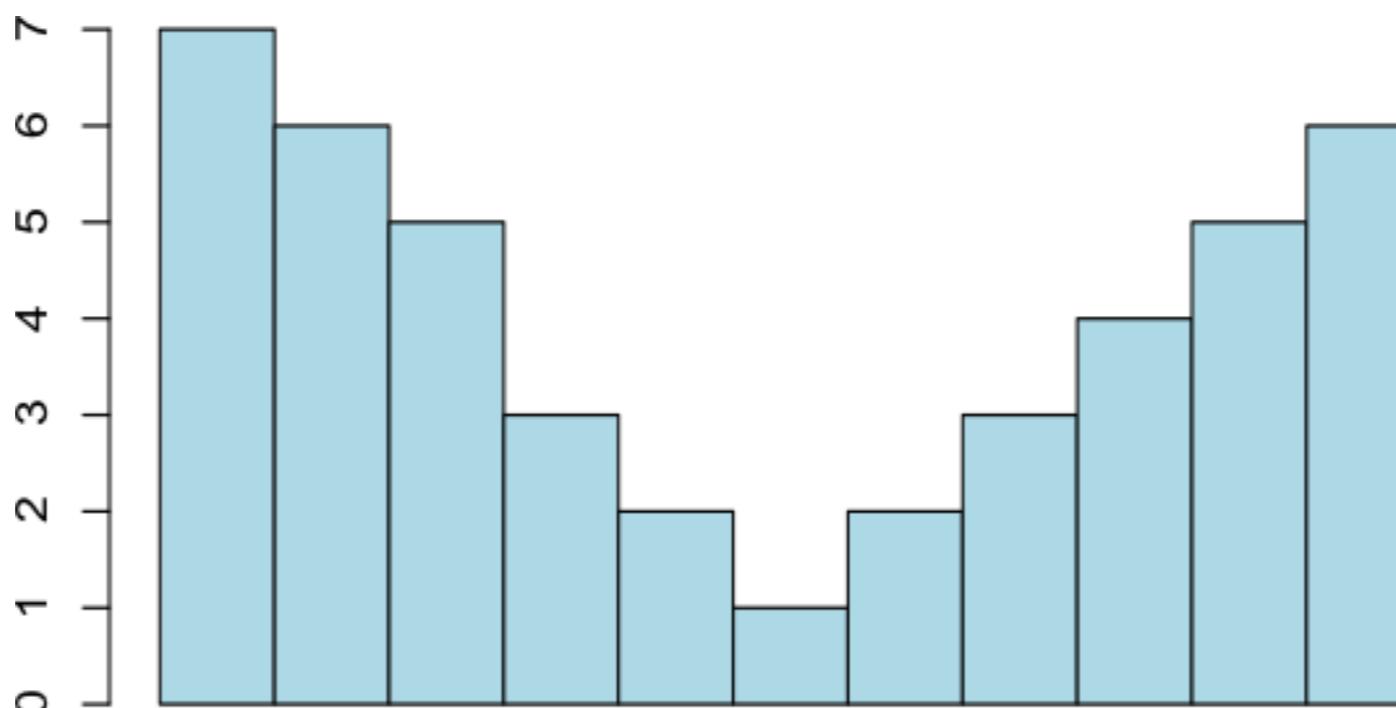


Flat

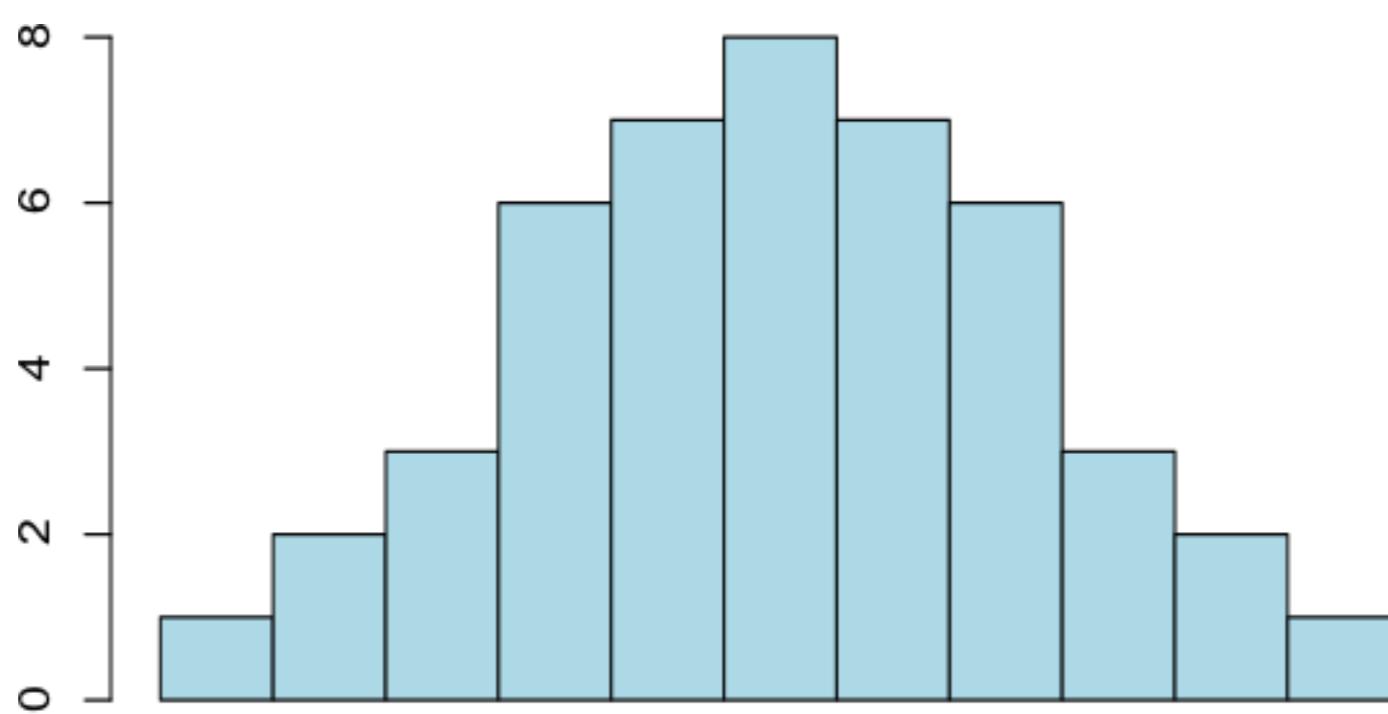




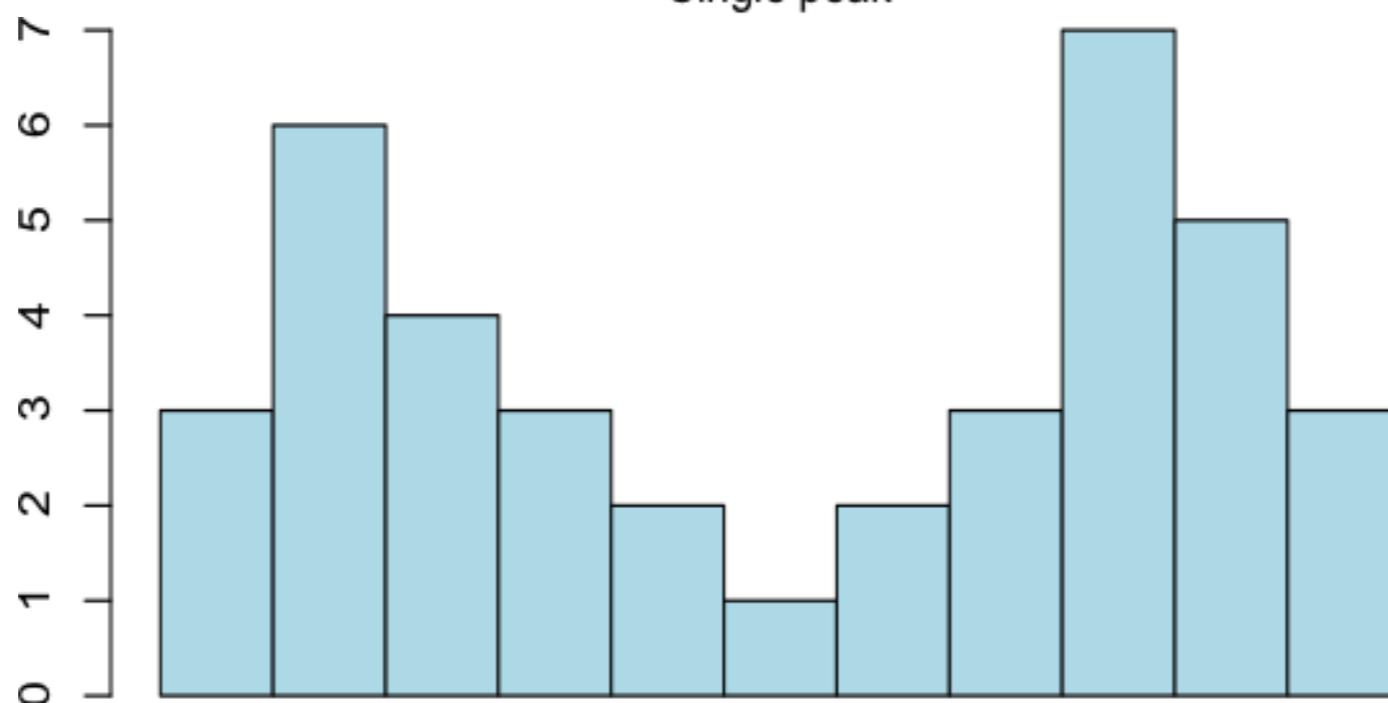
Upward



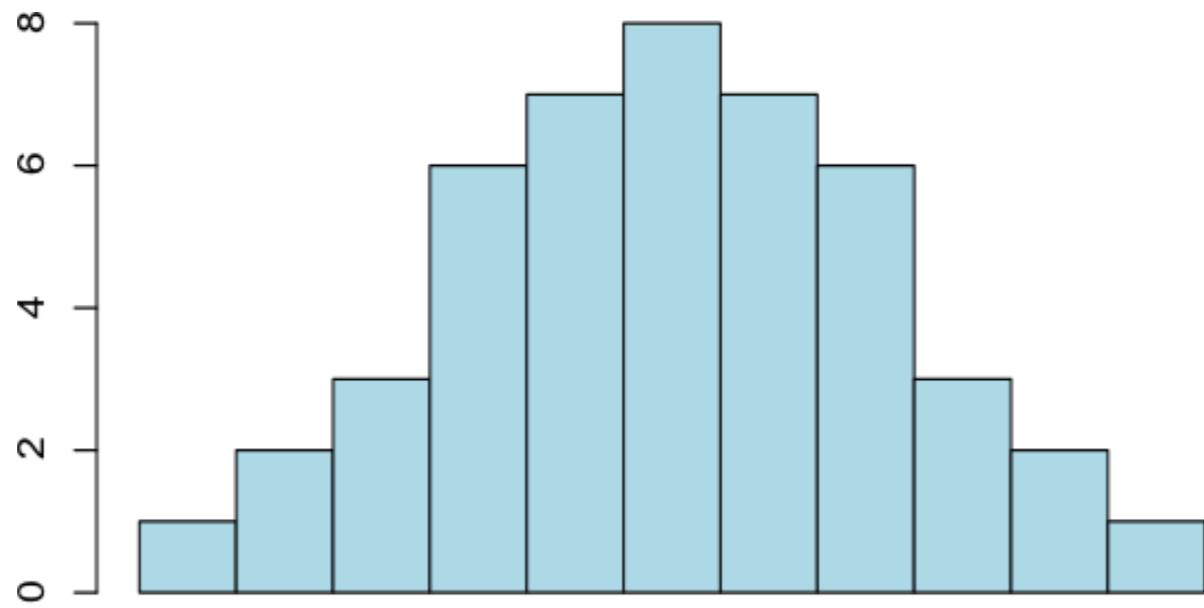
Downward



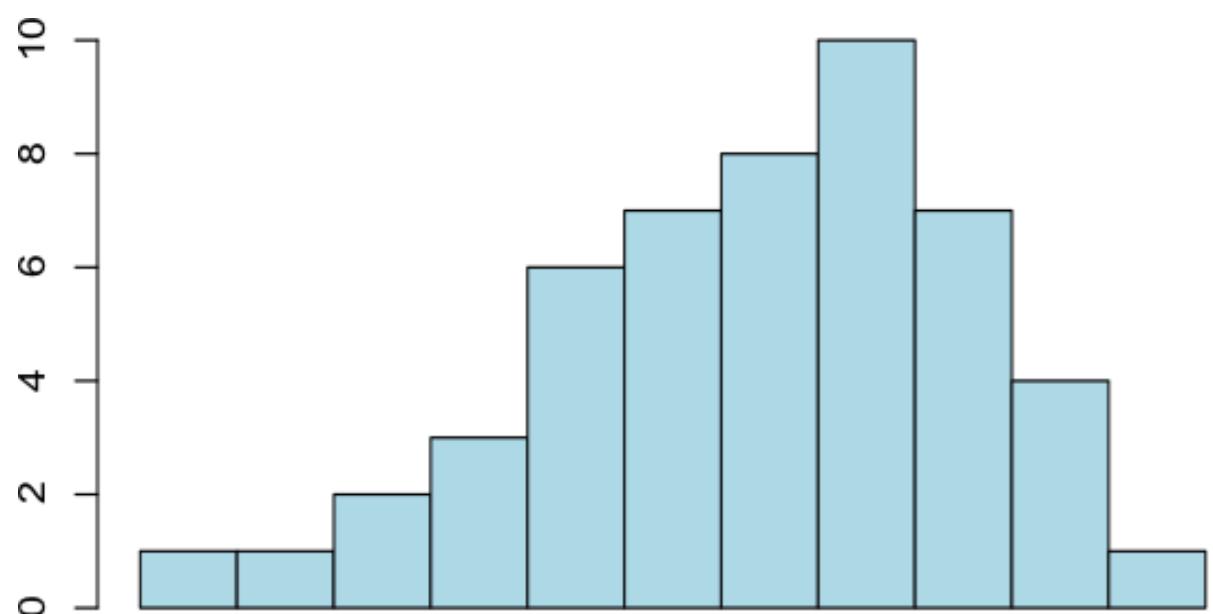
Single peak



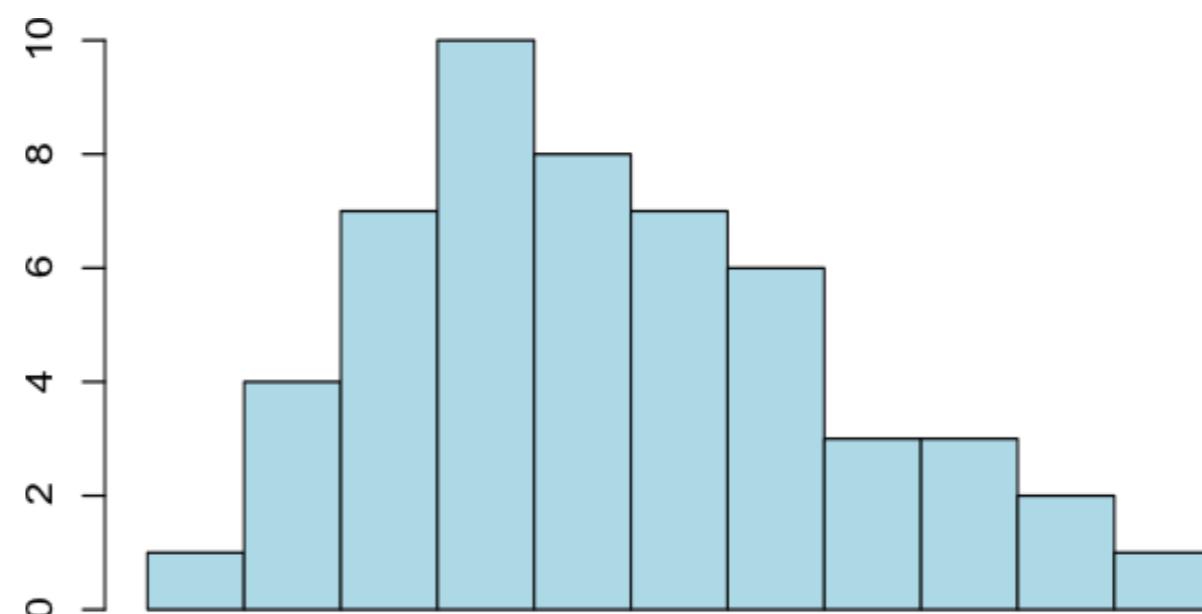
Multiple peaks



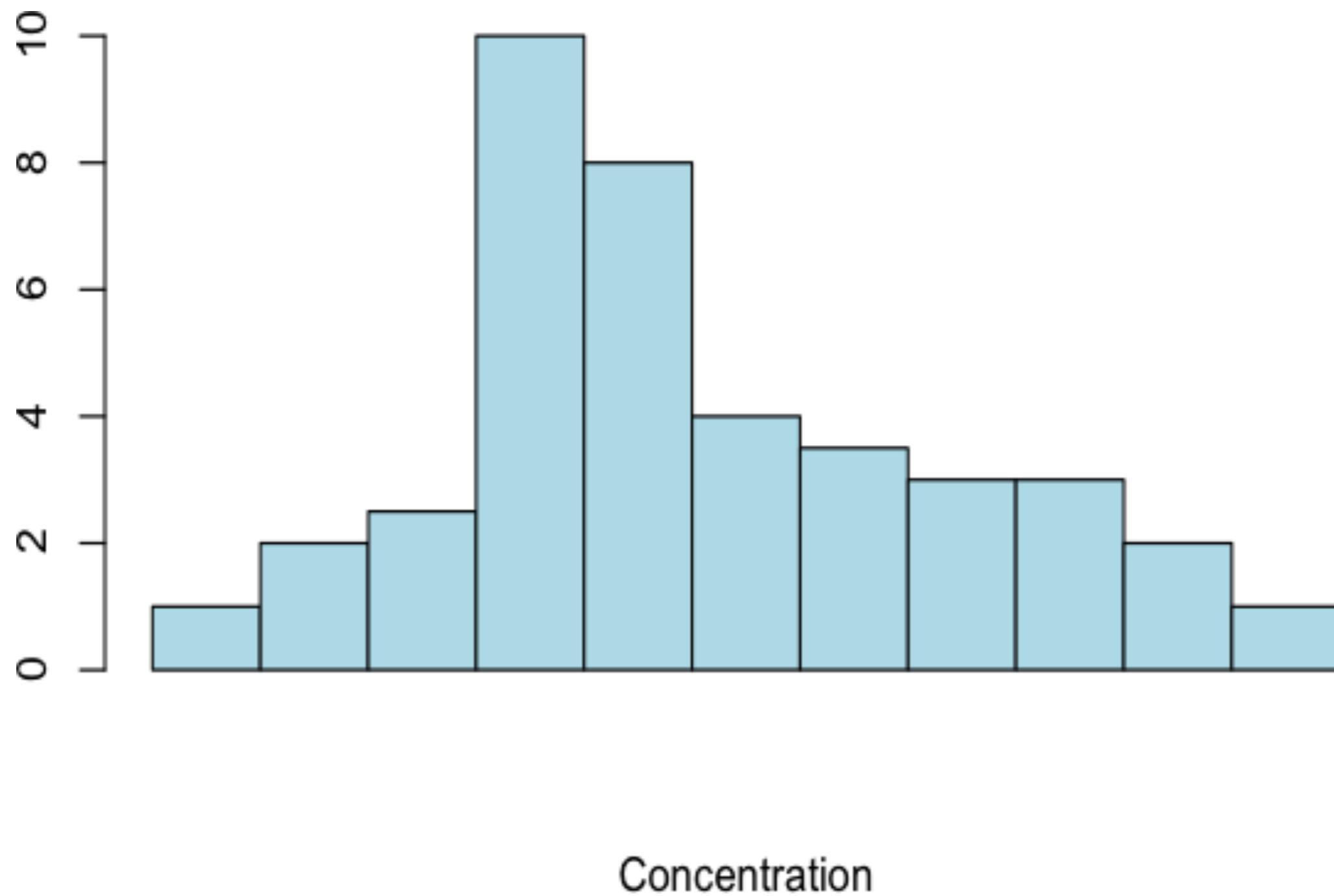
Symmetrical

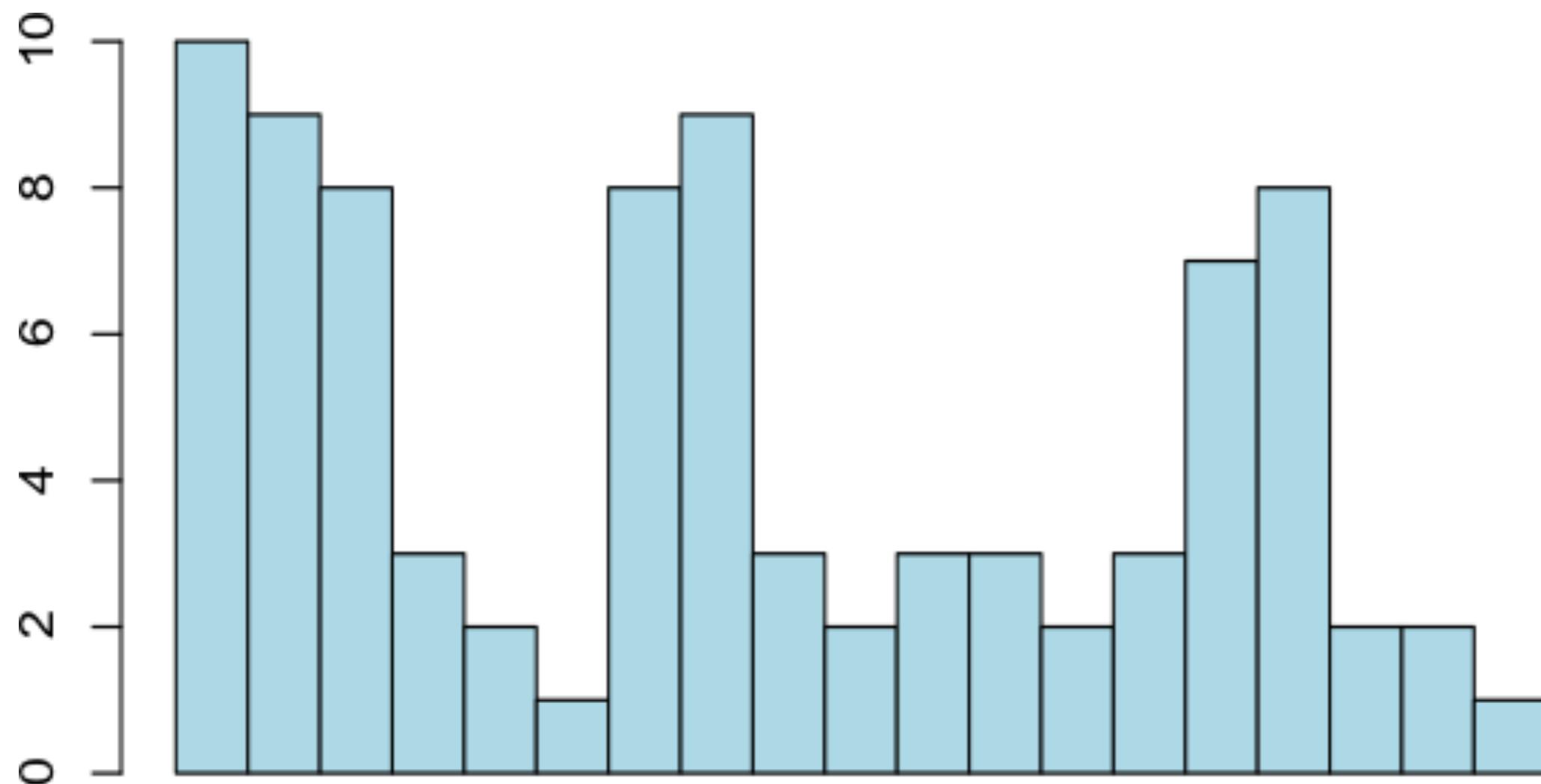


Skewed to the left

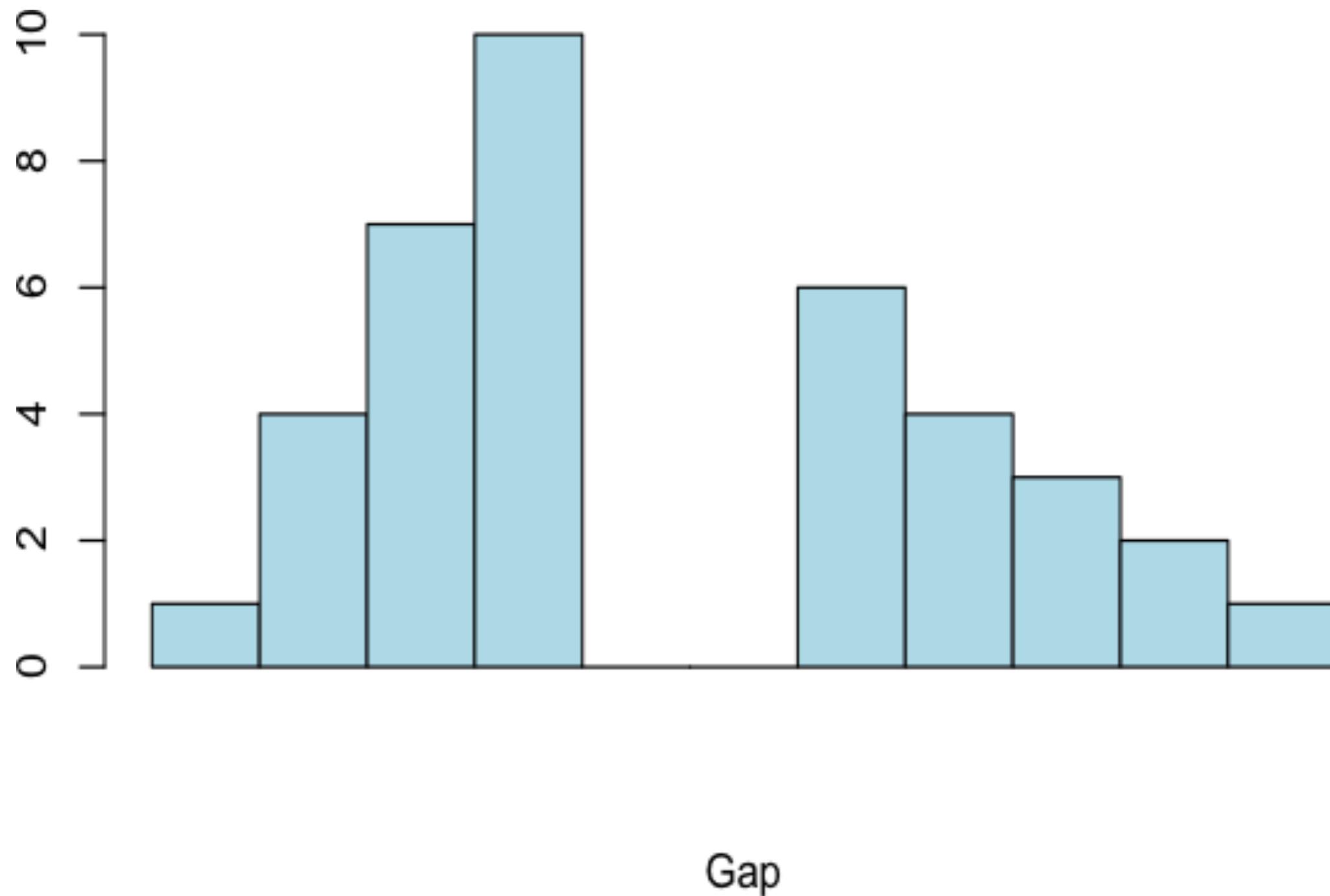


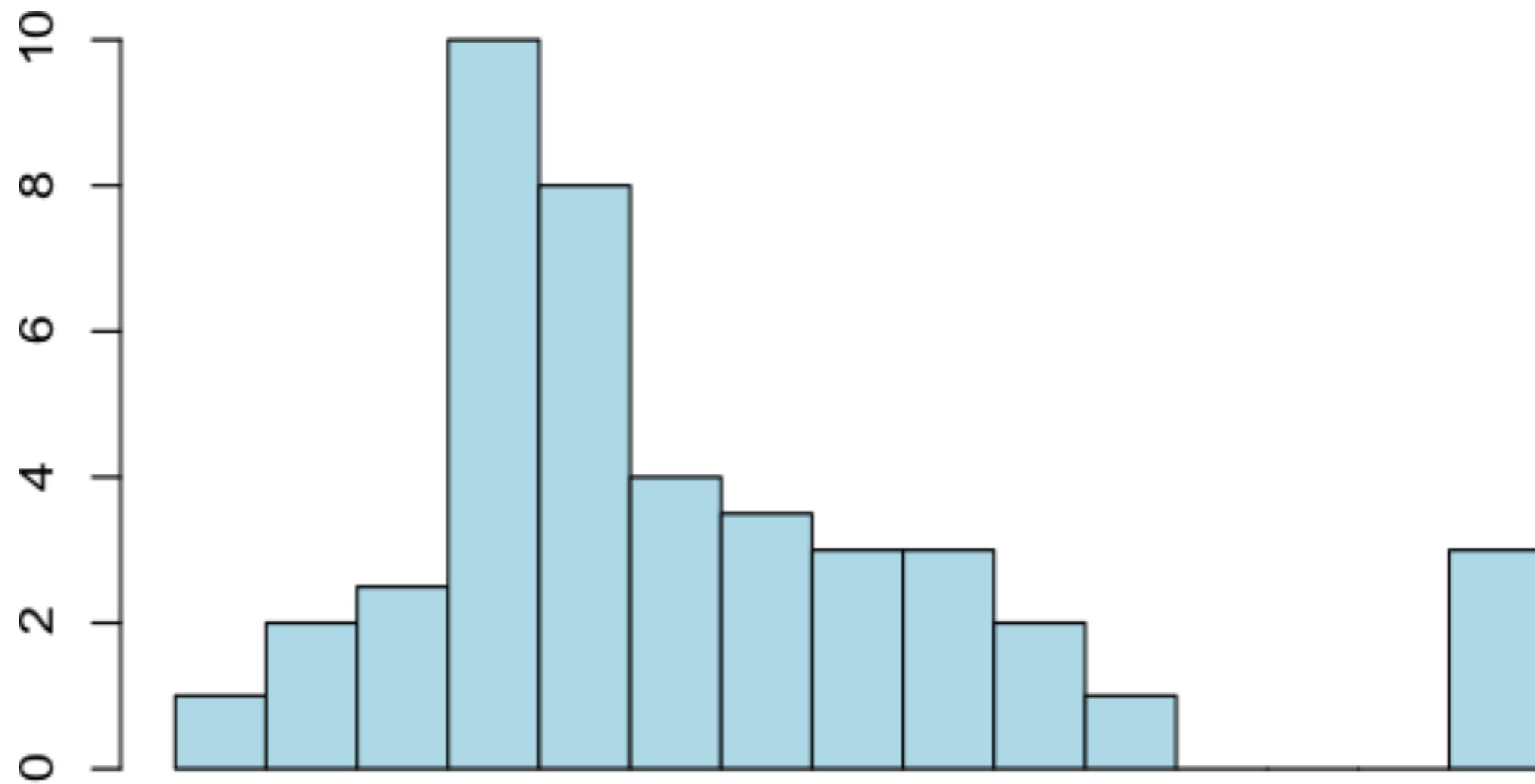
Skewed to the right





Concentration

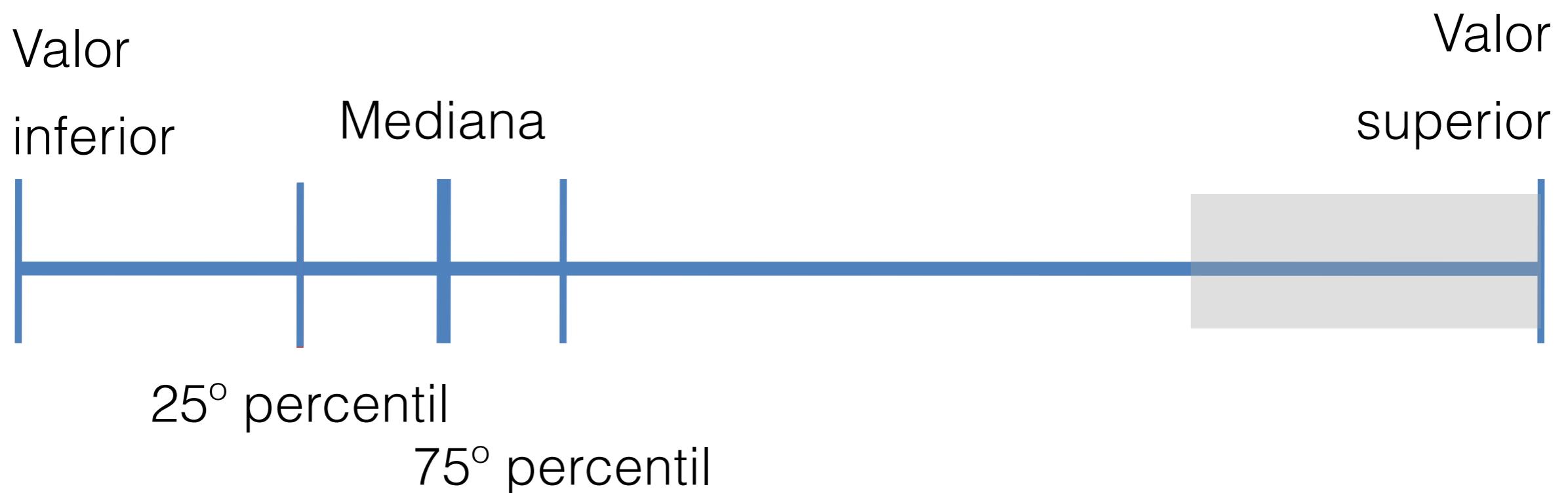




Outlier

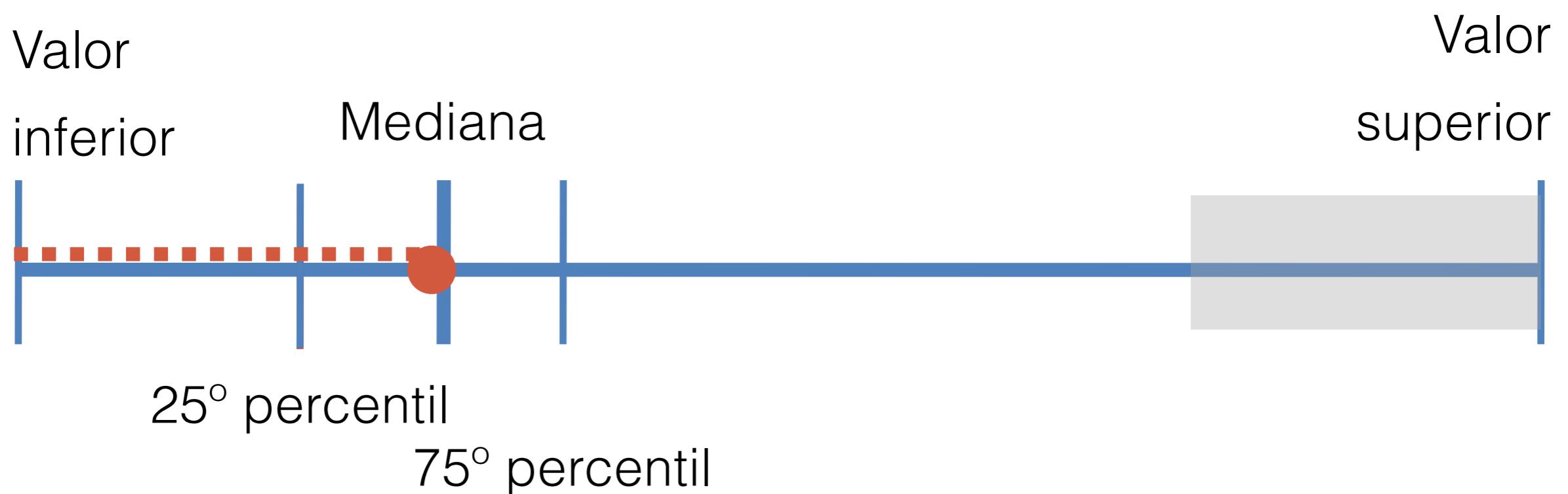
Regra simples para identificação de exceções

ponto médio entre 25° e 75° percentil multiplicado por 1,5 e subtraído e somado dos mesmos



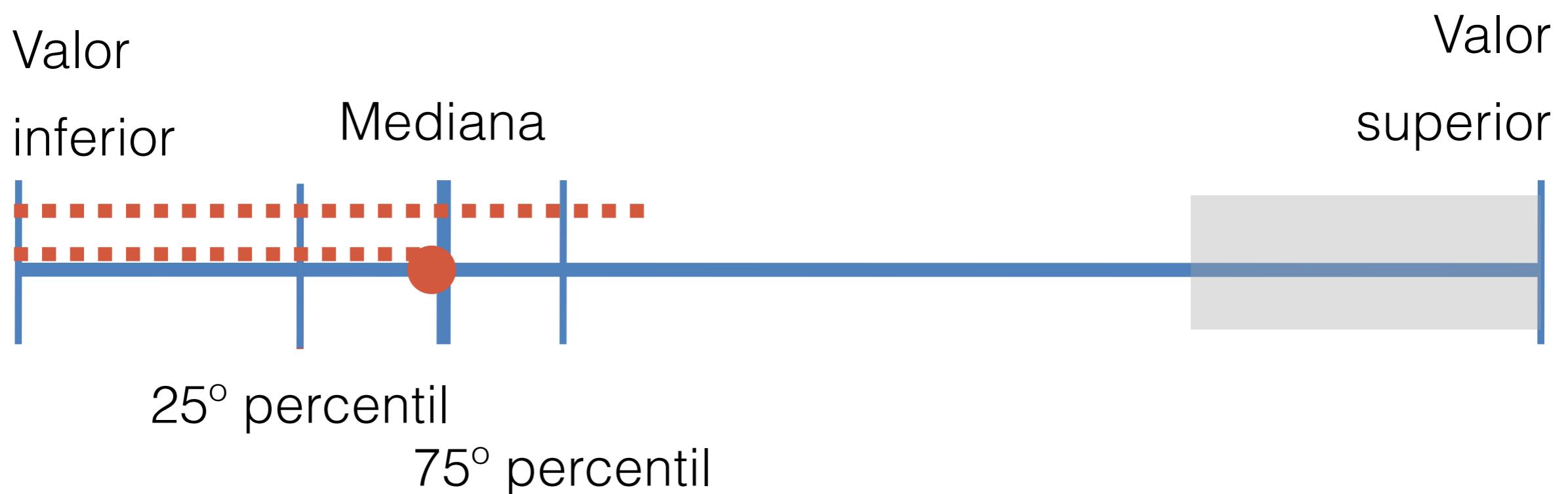
Regra simples para identificação de exceções

ponto médio entre 25° e 75° percentil multiplicado por 1,5 e subtraído e somado dos mesmos



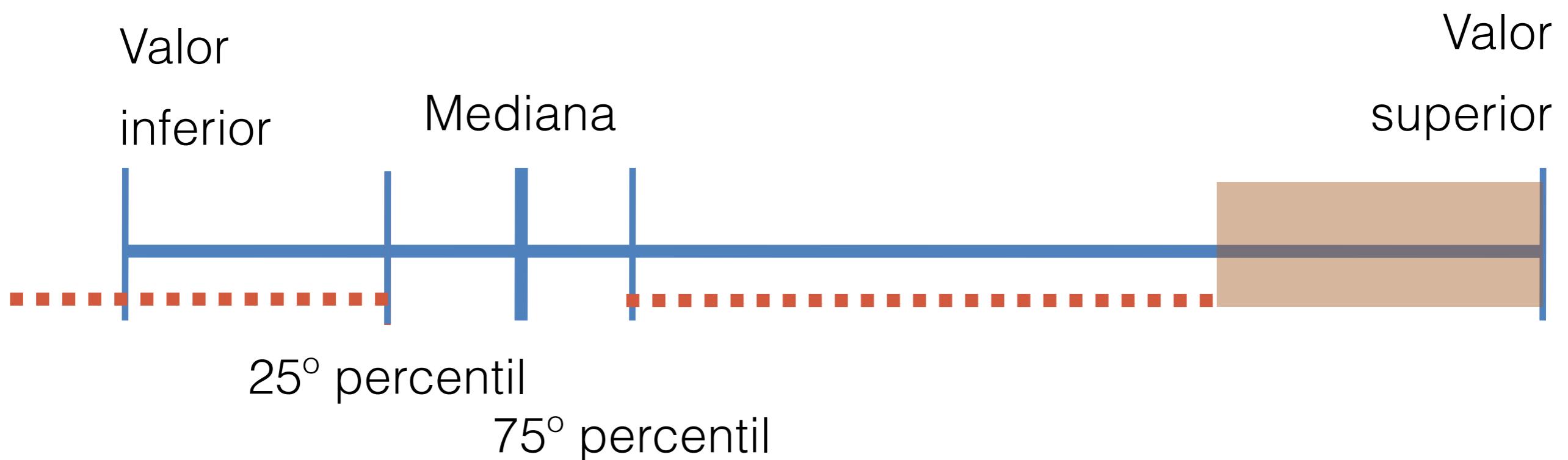
Regra simples para identificação de exceções

ponto médio entre 25° e 75° percentil multiplicado por 1,5 e subtraído e somado dos mesmos

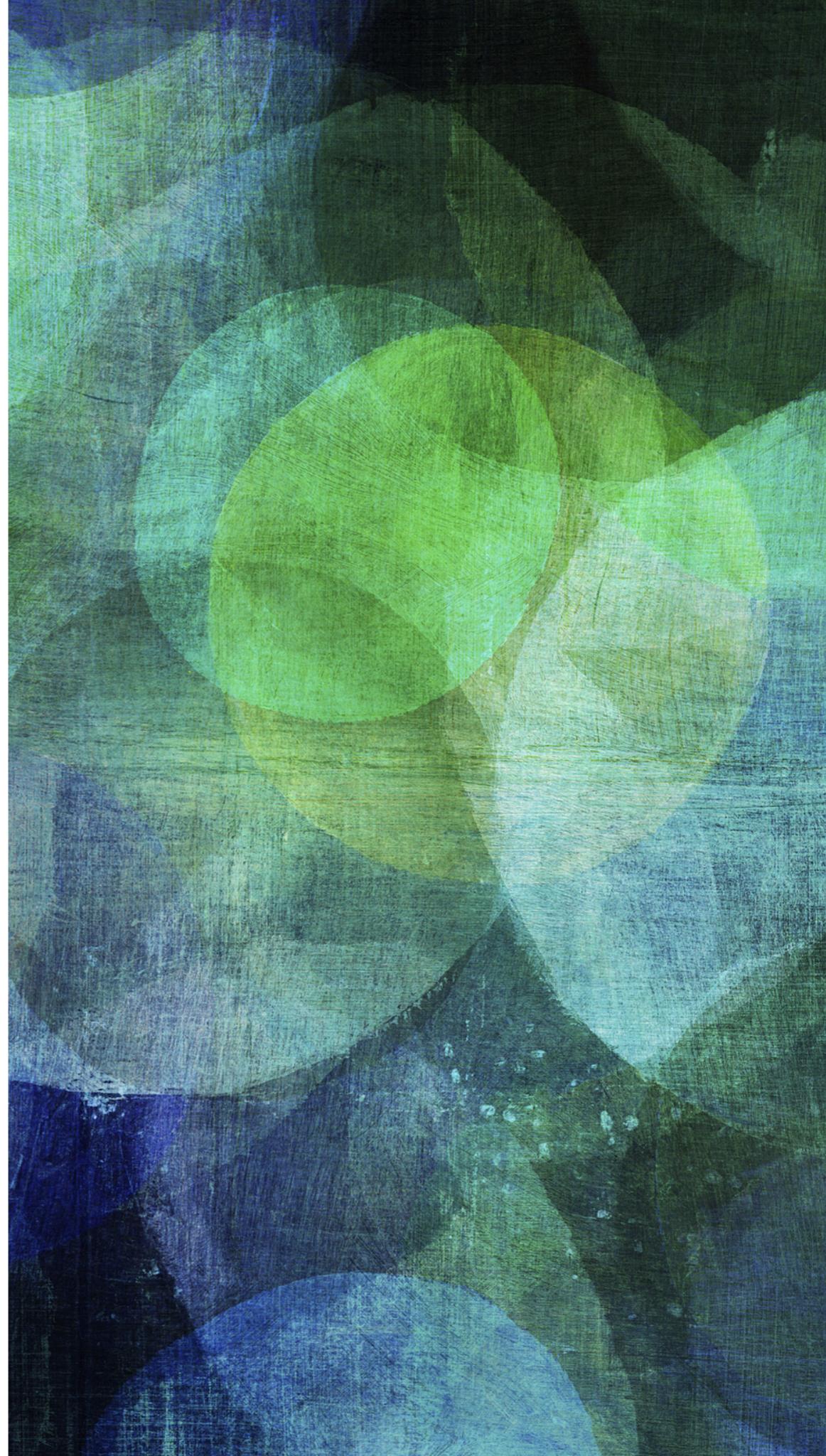


Regra simples para identificação de exceções

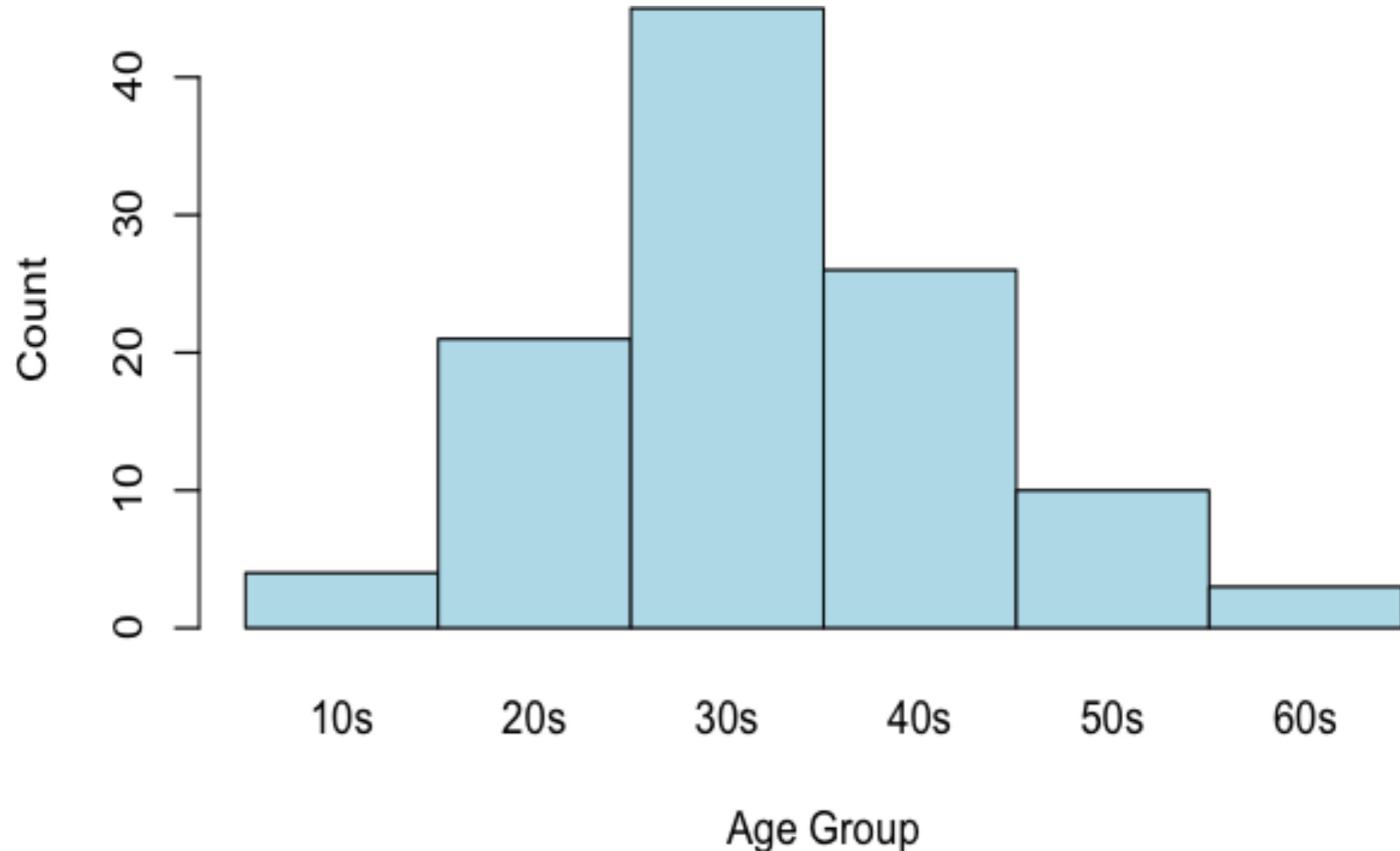
ponto médio entre 25° e 75° percentil multiplicado por 1,5 e subtraído e somado dos mesmos



REPRESENTAÇÕES VISUAIS

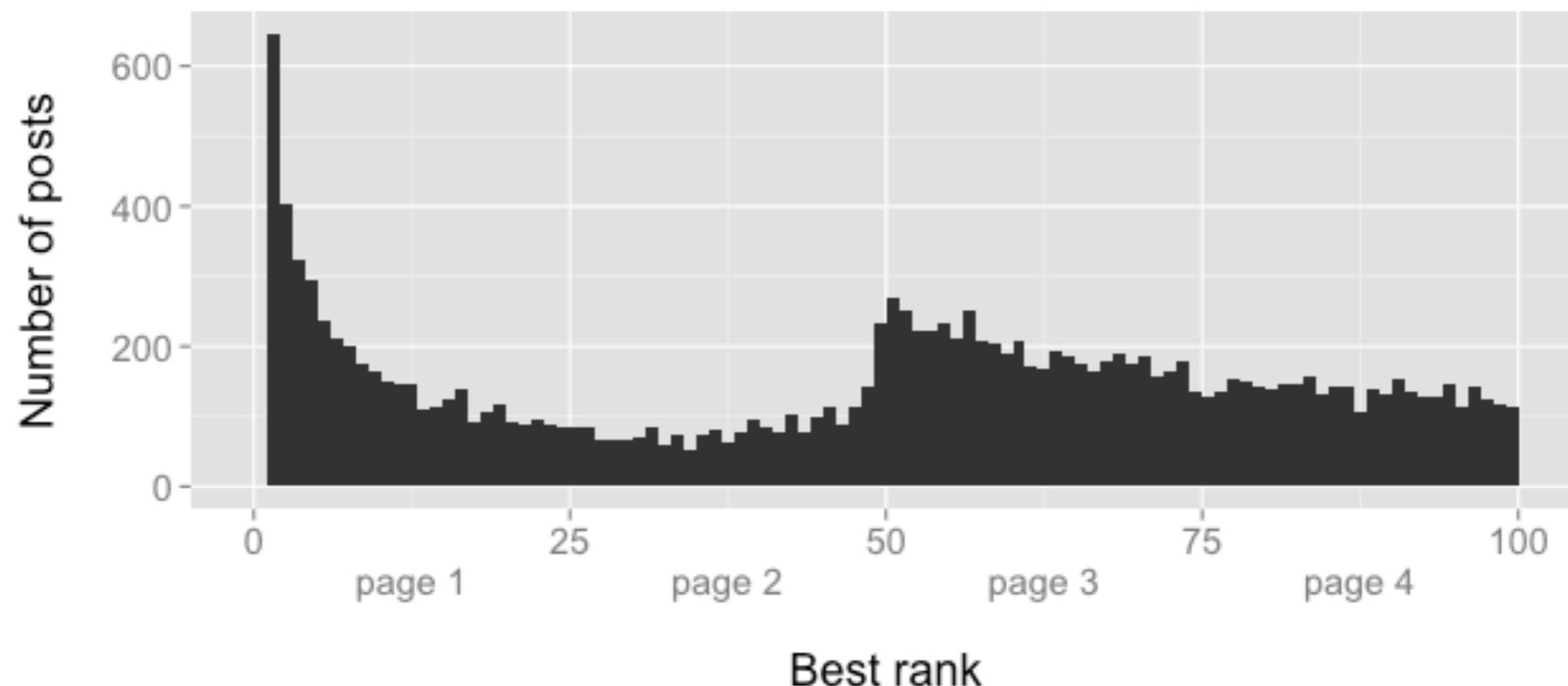


HISTOGRAMA



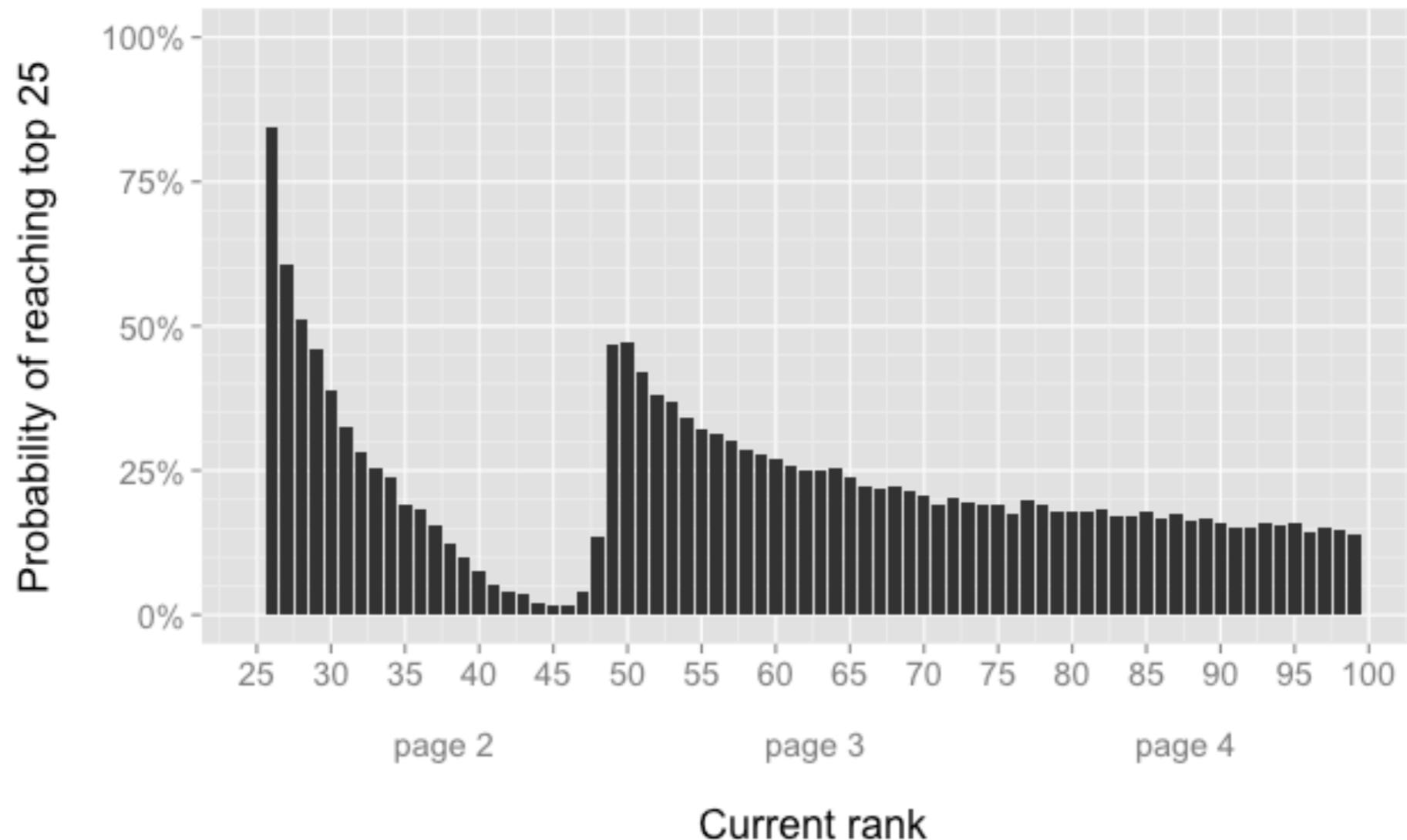
HISTOGRAMA

Distribution of best rank for 14979 posts observed in reddit top 100 between Sep 18 and Oct 31, 2014

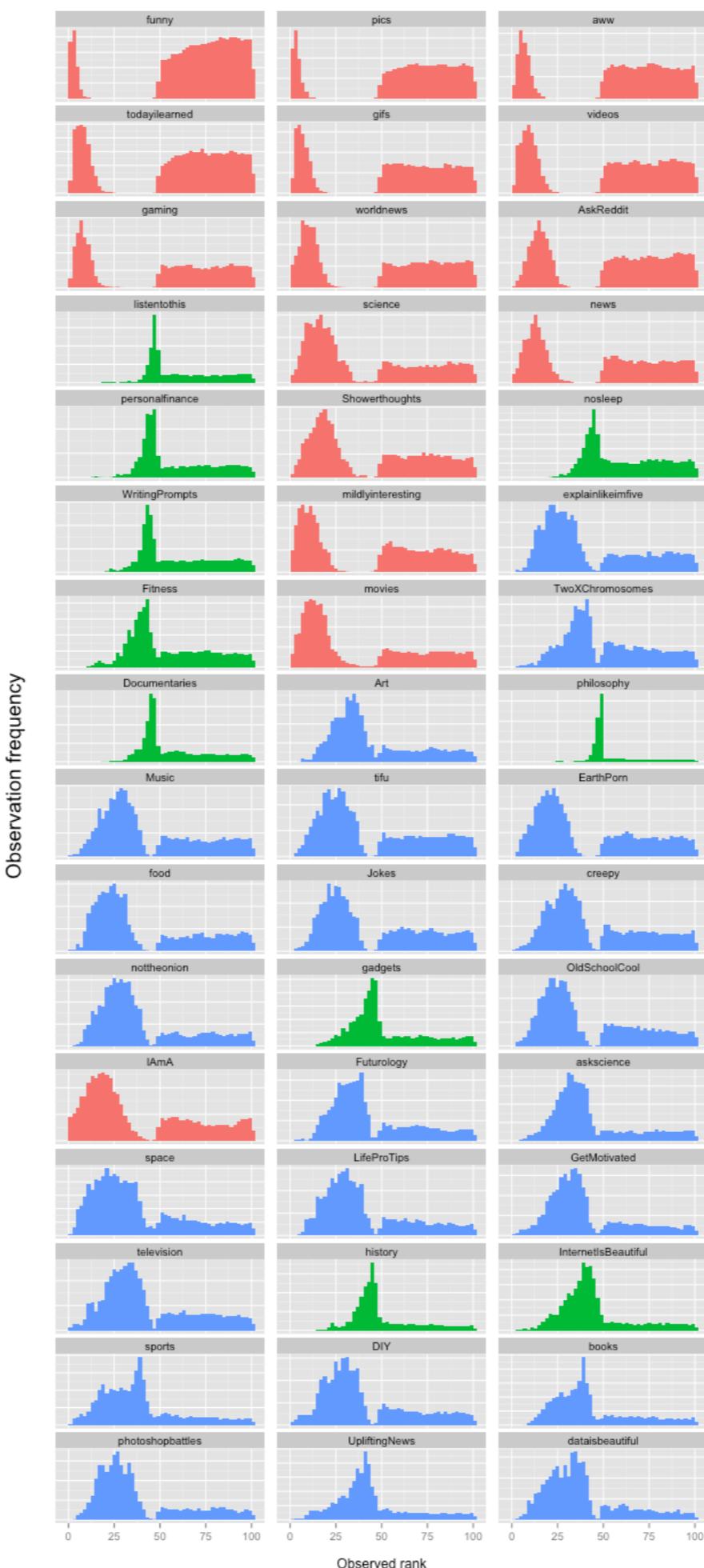


HISTOGRAMA

Conditional probability of a reddit post reaching the logged-out homepage, given that it's in the top 100 and its rank is not declining



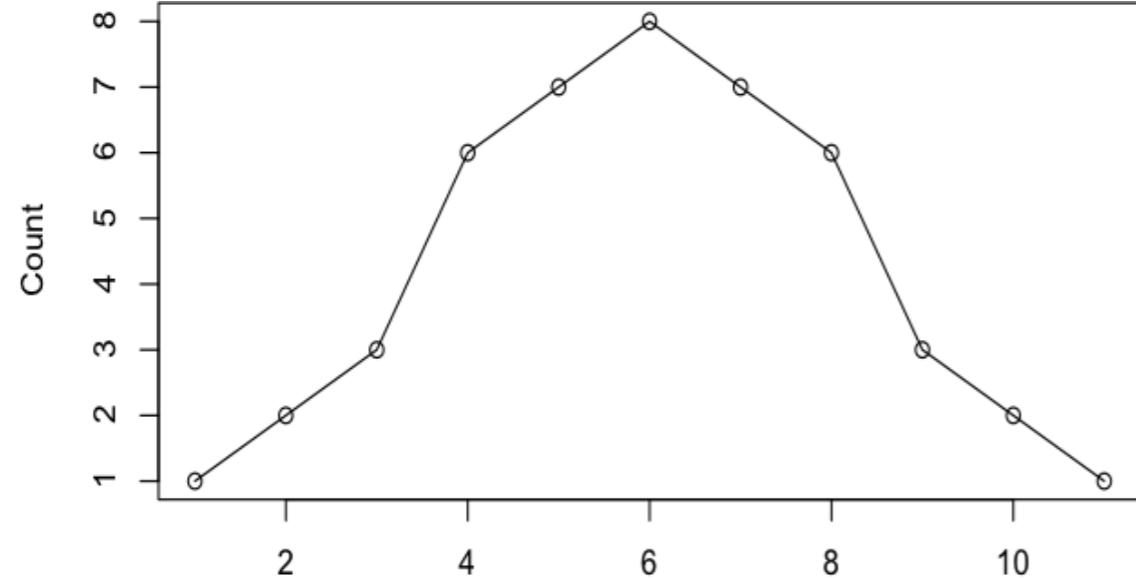
Observed ranks by subreddit



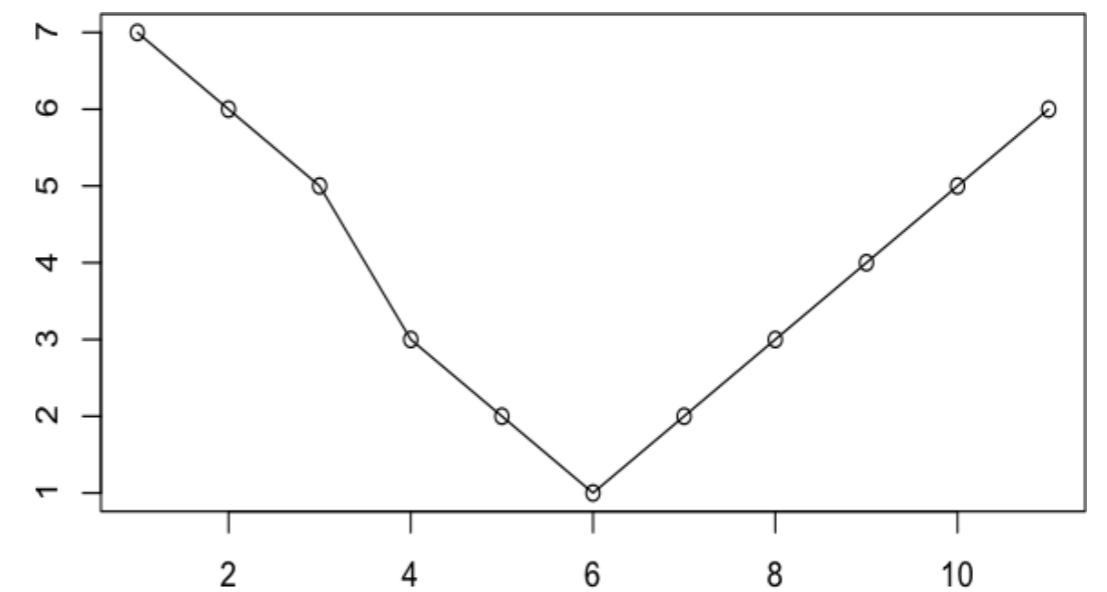
POLÍGONO DE FREQUÊNCIA



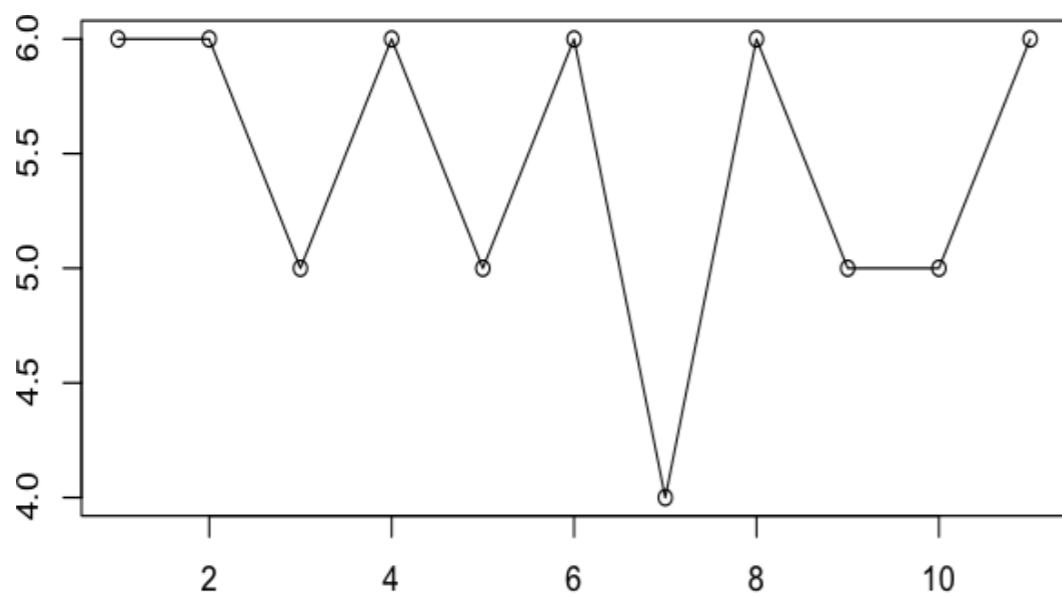
MÚLTIPLOS POLÍGONO DE FREQUÊNCIA



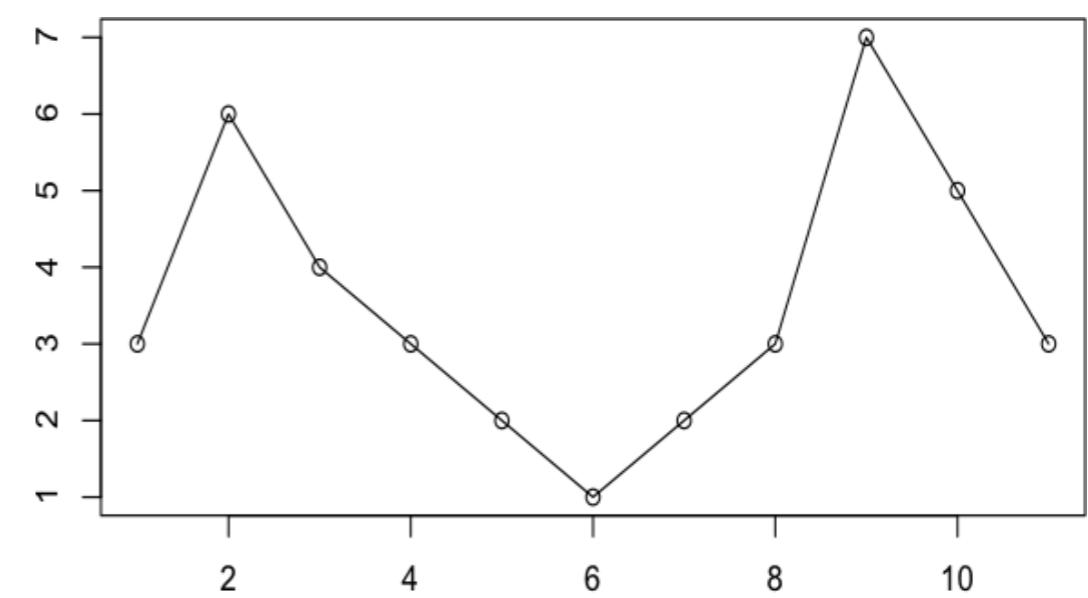
A



C

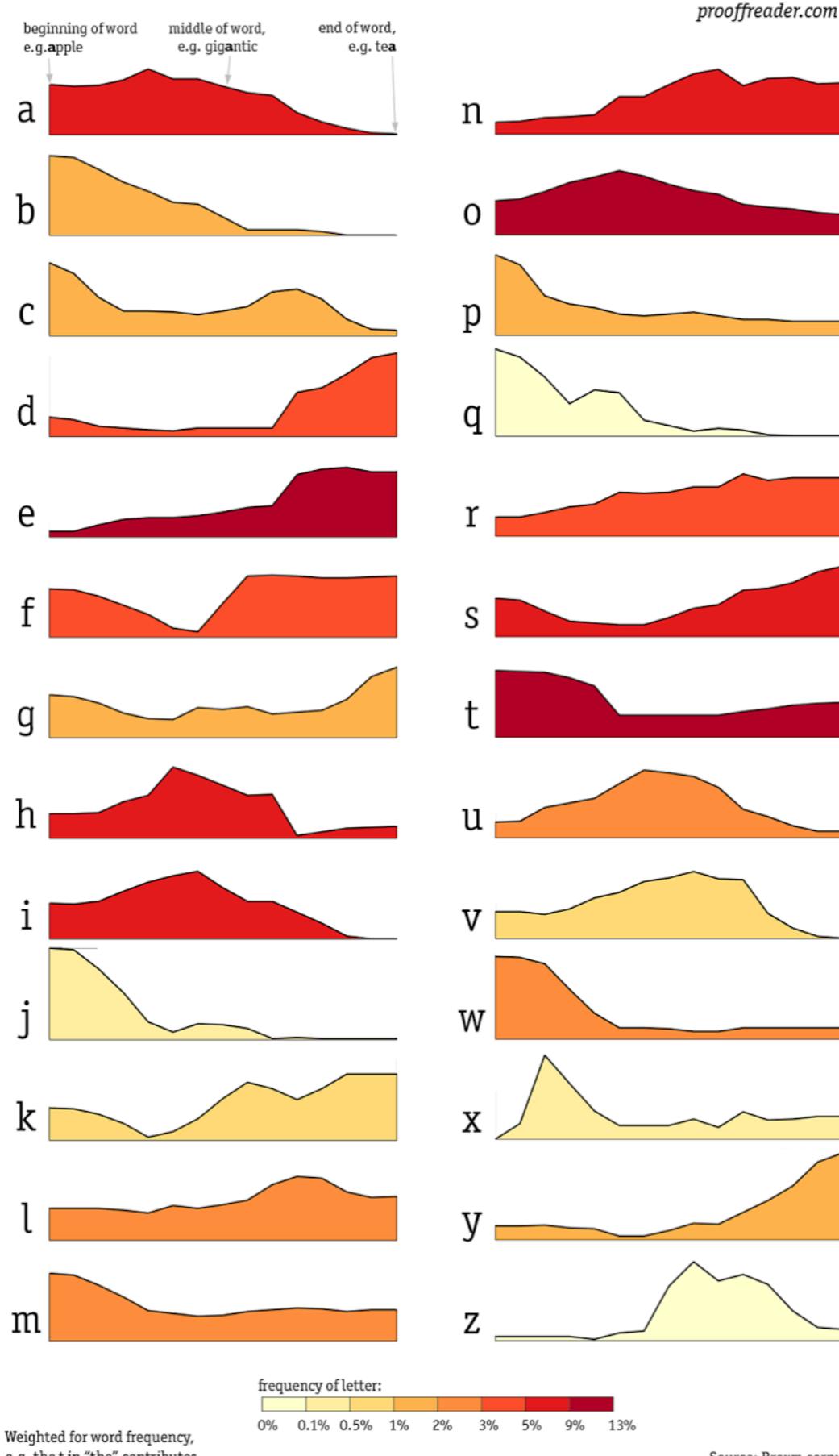


B



D

Distribution of English letters toward beginning, middle and end of words



STEM-AND-LEAF (RAMO E FOLHAS)

1. 8 9 9 9

2. 0 0 1 1 3 3 3 4 4 4 5 6 6 6 6 6 7 7 8 8 9 9 9 9 9

3. 0 0 0 0 1 1 2 2 2 2 2 2 3 3 3 3 3 4 4 4 5 5 5 6 7 8 9

4. 0 0 0 0 1 1 2 2 3 3 4 4 5 5 5 6 6 6 7 8 9

5. 0 0 1 3 5 7 9

6. 1 4 8

Usar no máximo 2 valores no stem e 3 valores no leaf

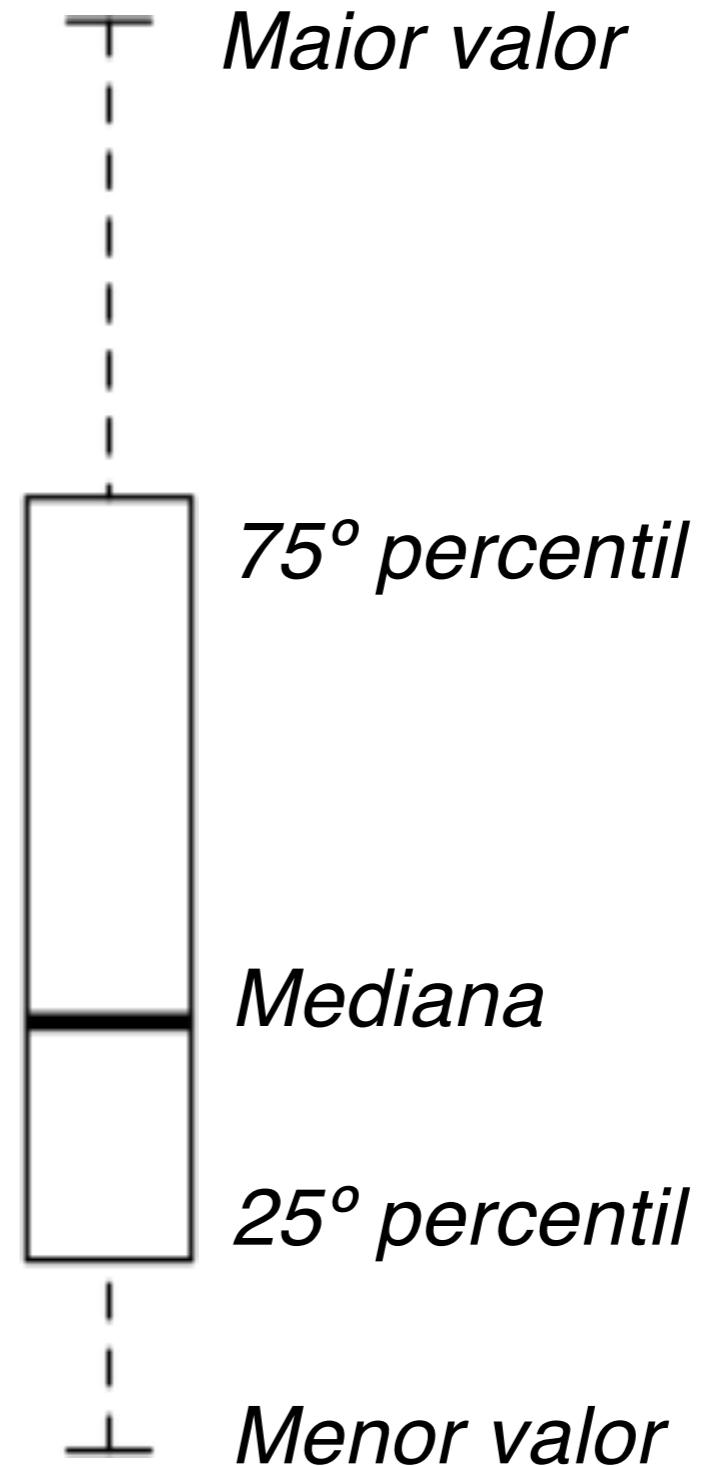
STEM-AND-LEAF (RAMO E FOLHAS)



BOX PLOT

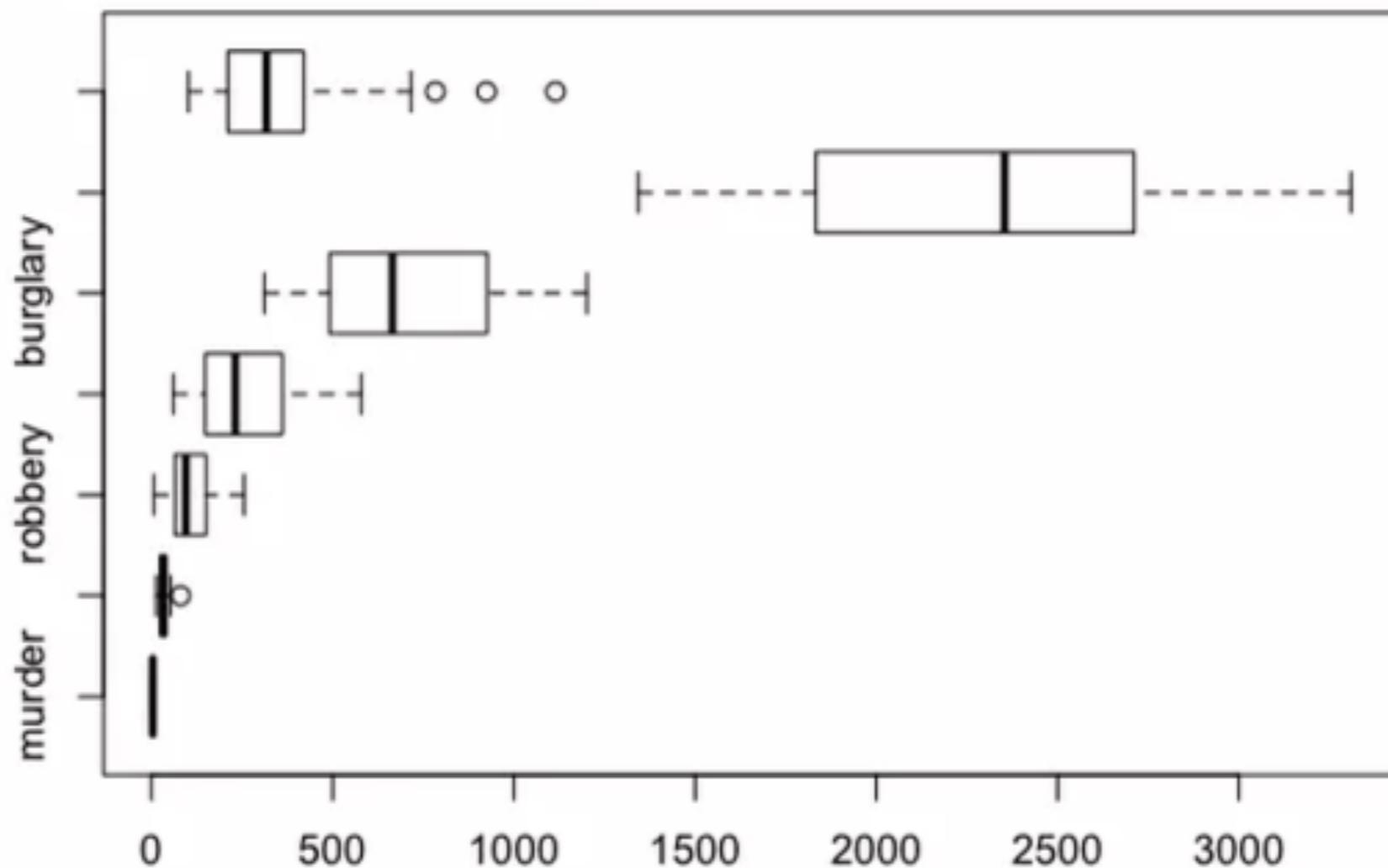
*Espalhamento
de 100%
dos valores*

*Espalhamento
de 50%
dos valores*

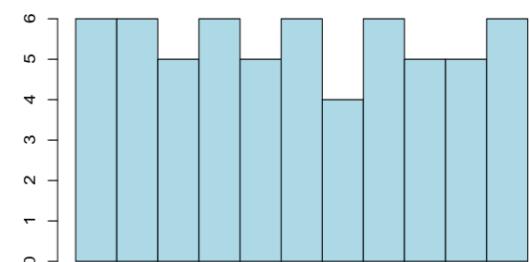
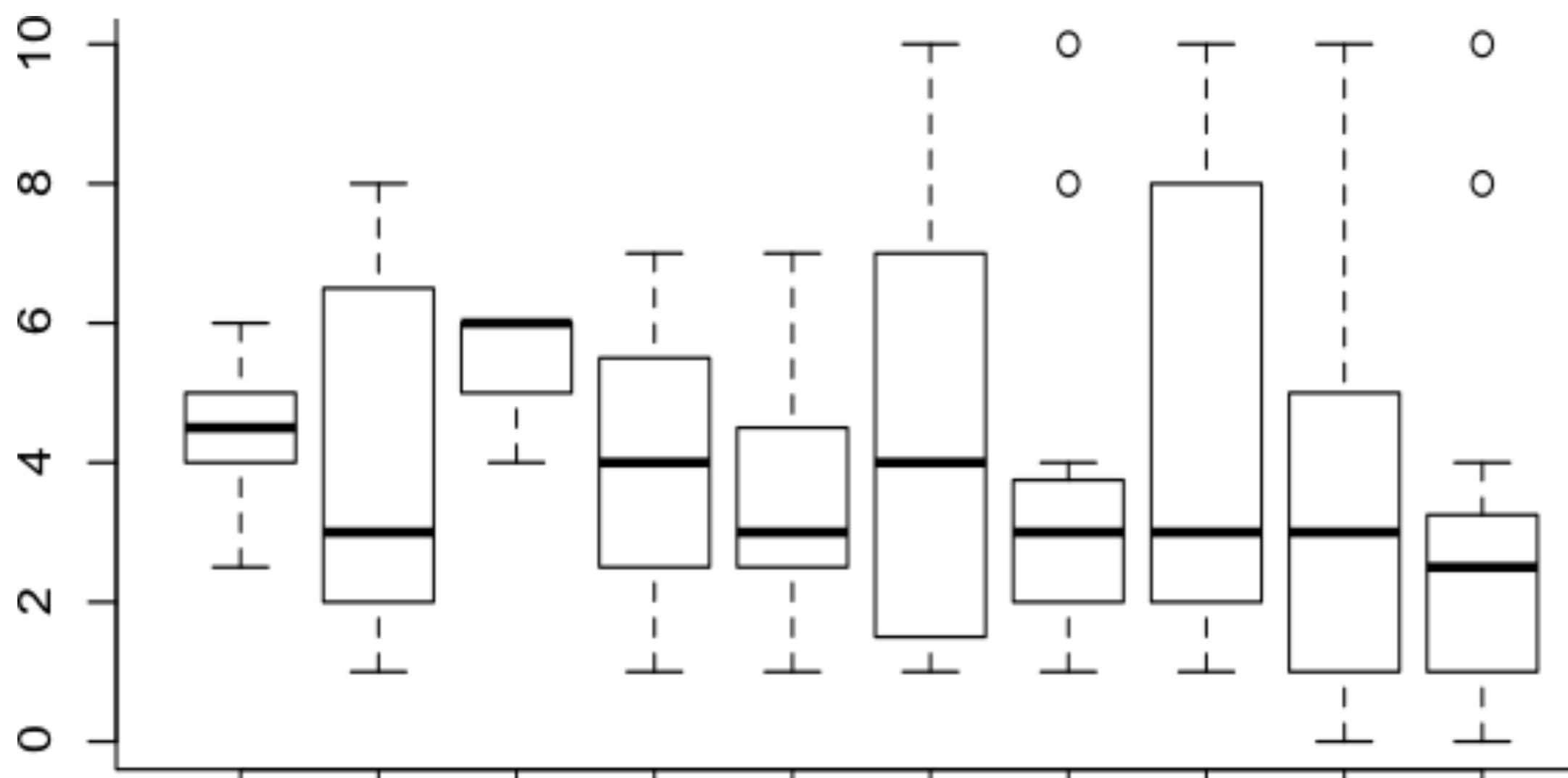


BOX PLOT

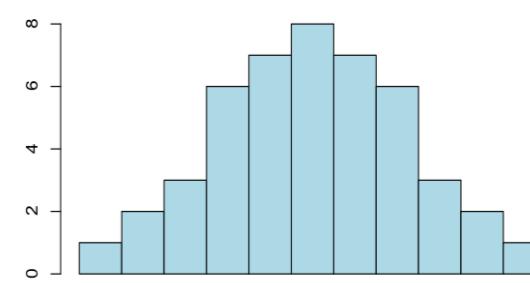
Crime Rates in US



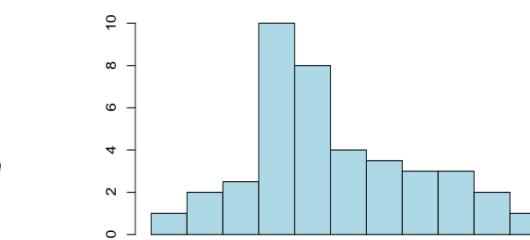
BOX PLOT



3°



4°



7°

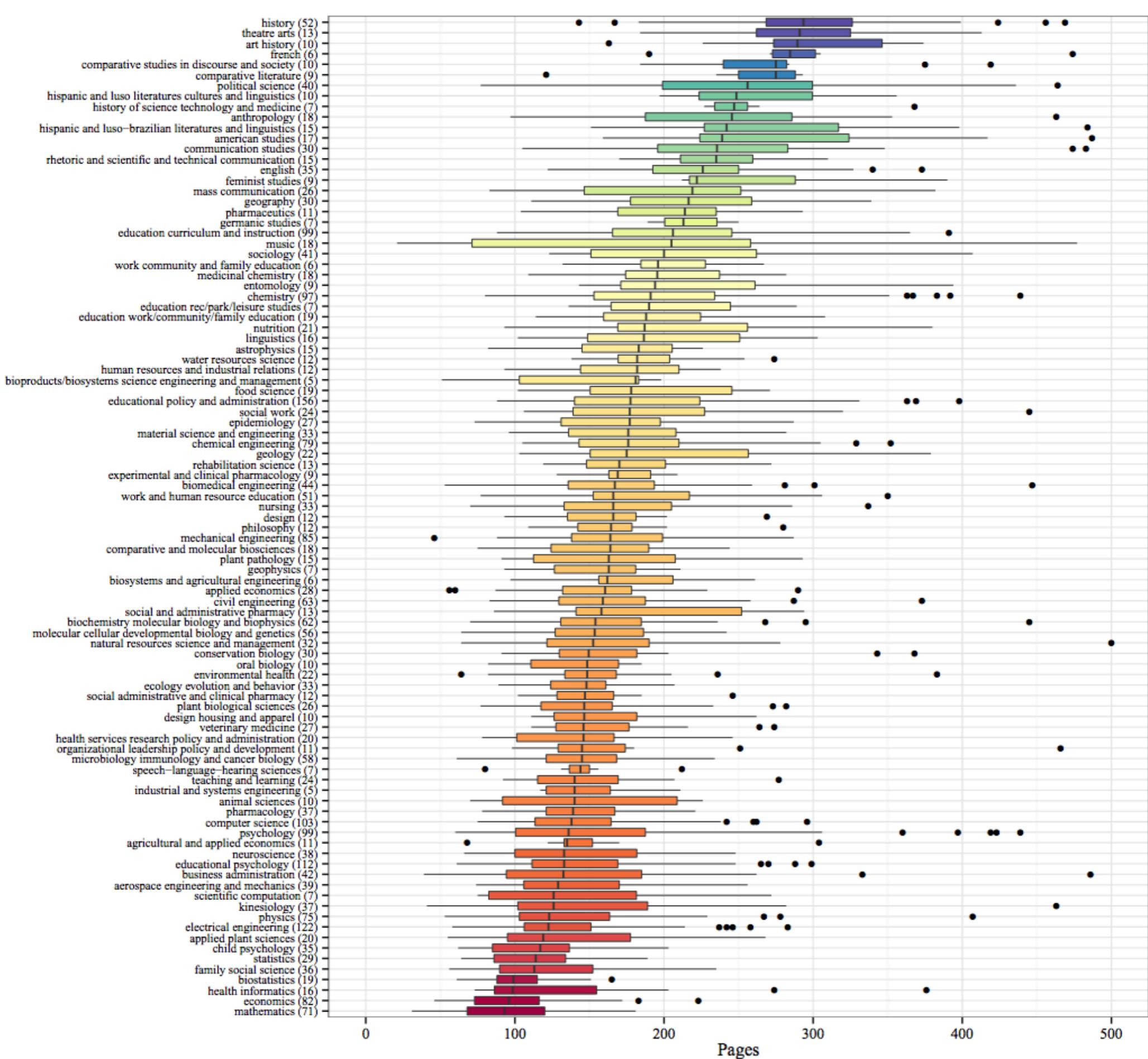


Fig: Summary of page lengths of doctoral dissertations by major, sorted and color-coded by median. Boxes represent the median, 25th and 75th percentiles, 1.5 times the interquartile range as whiskers, and outliers beyond the whiskers. Number of records for each major are in parentheses.

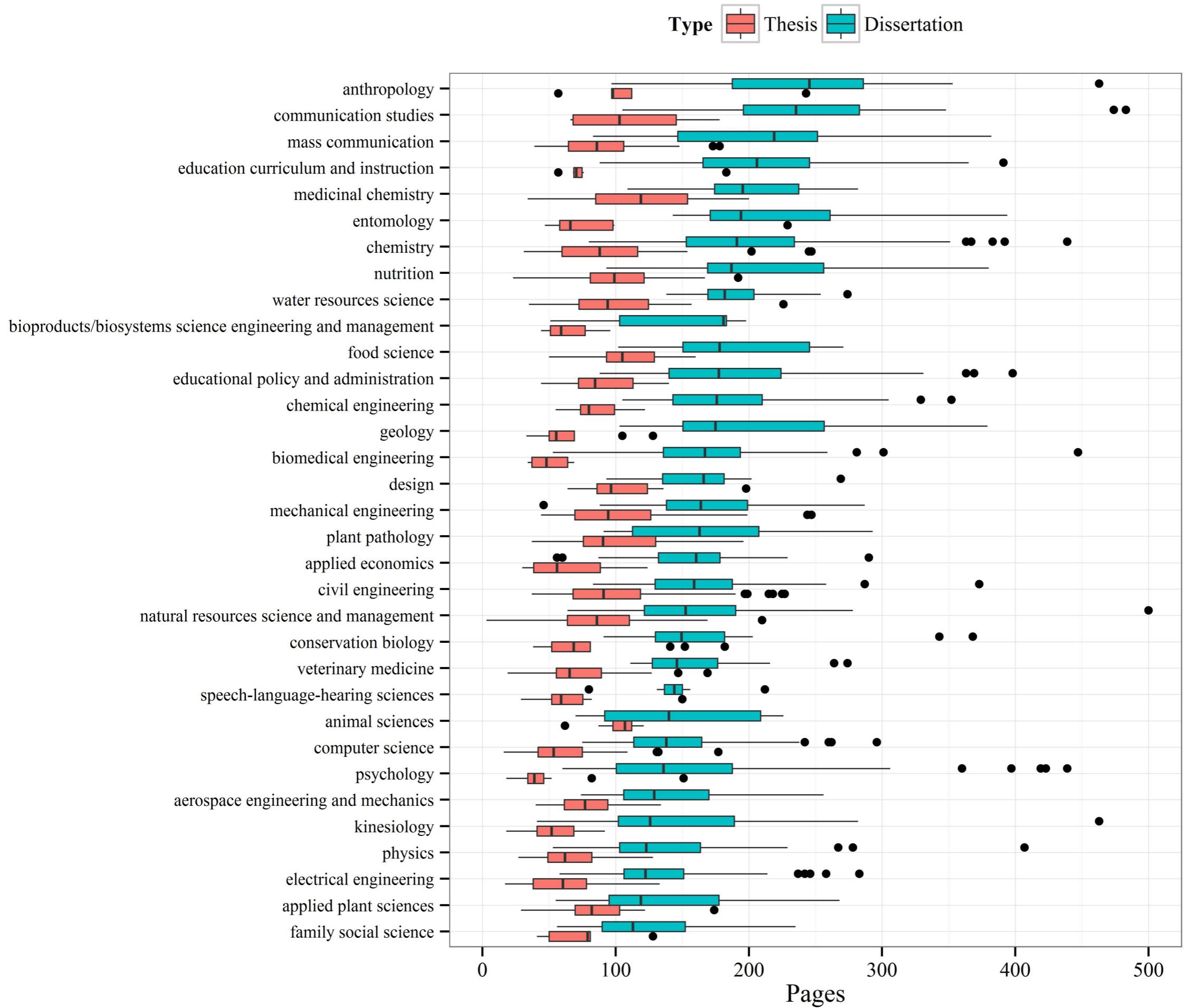


Fig: Comparison of dissertation and thesis page lengths for majors having both degree programs in the database. Boxes represent the median, 25th and 75th percentiles, 1.5 times the interquartile range as whiskers, and outliers beyond the whiskers.

Ranking the Ivies

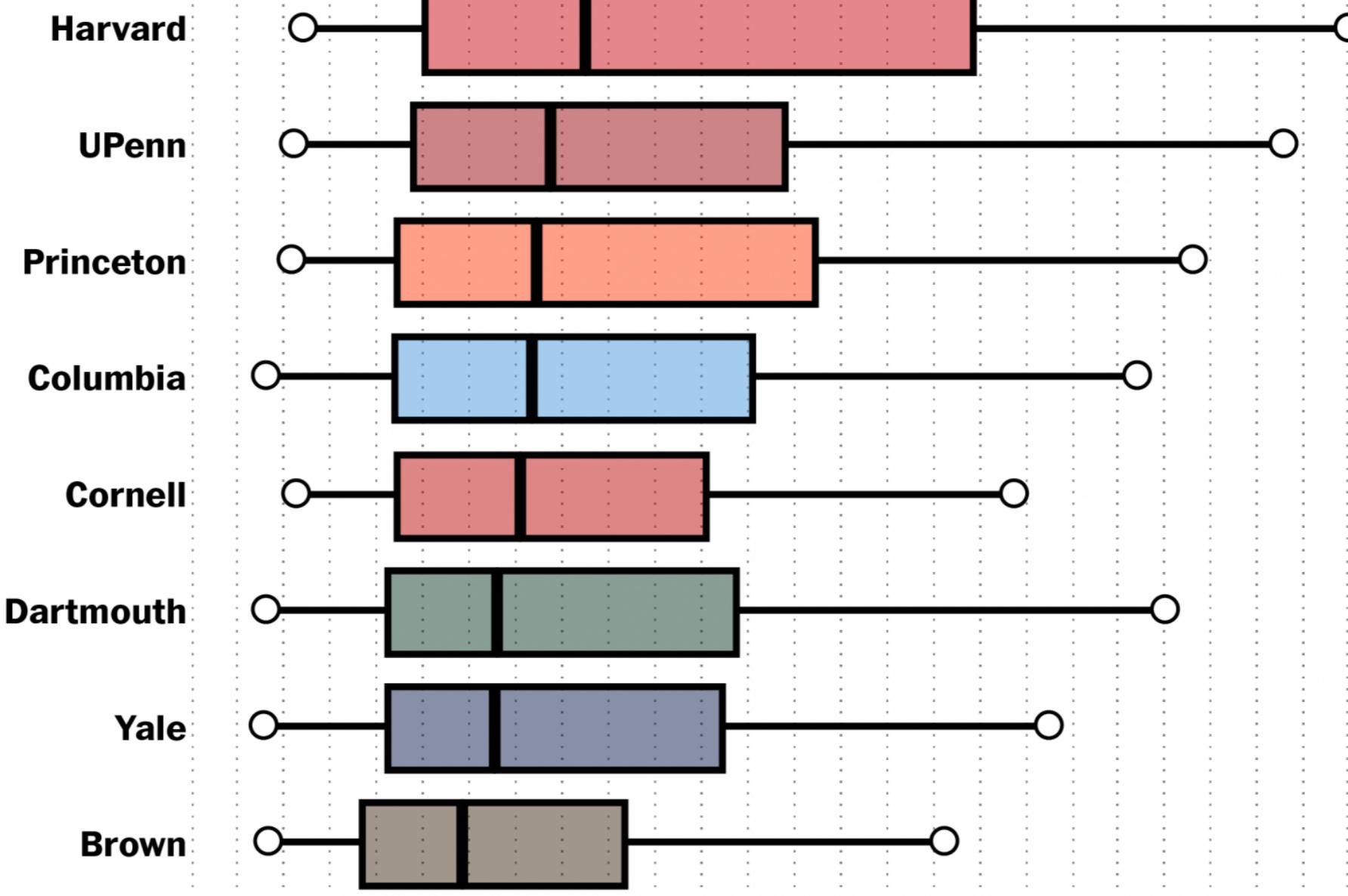
Annual earnings distributions, 10 years after starting school

How to read



10th percentile 25th percentile **Median annual earnings** 75th percentile 90th percentile

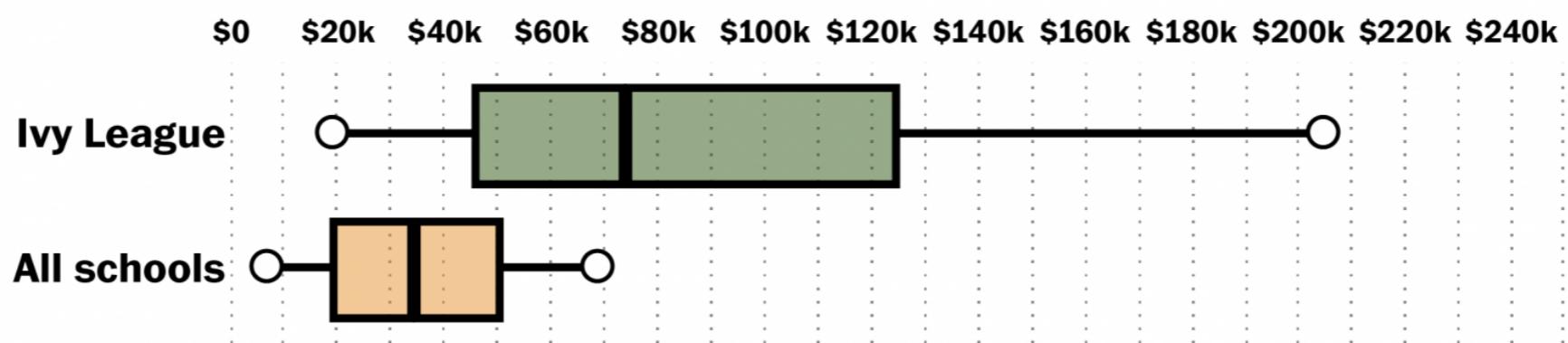
\$0 \$20k \$40k \$60k \$80k \$100k \$120k \$140k \$160k \$180k \$200k \$220k \$240k



Ivy League educations pay off

Annual earnings distributions, 10 years after starting school

How to read

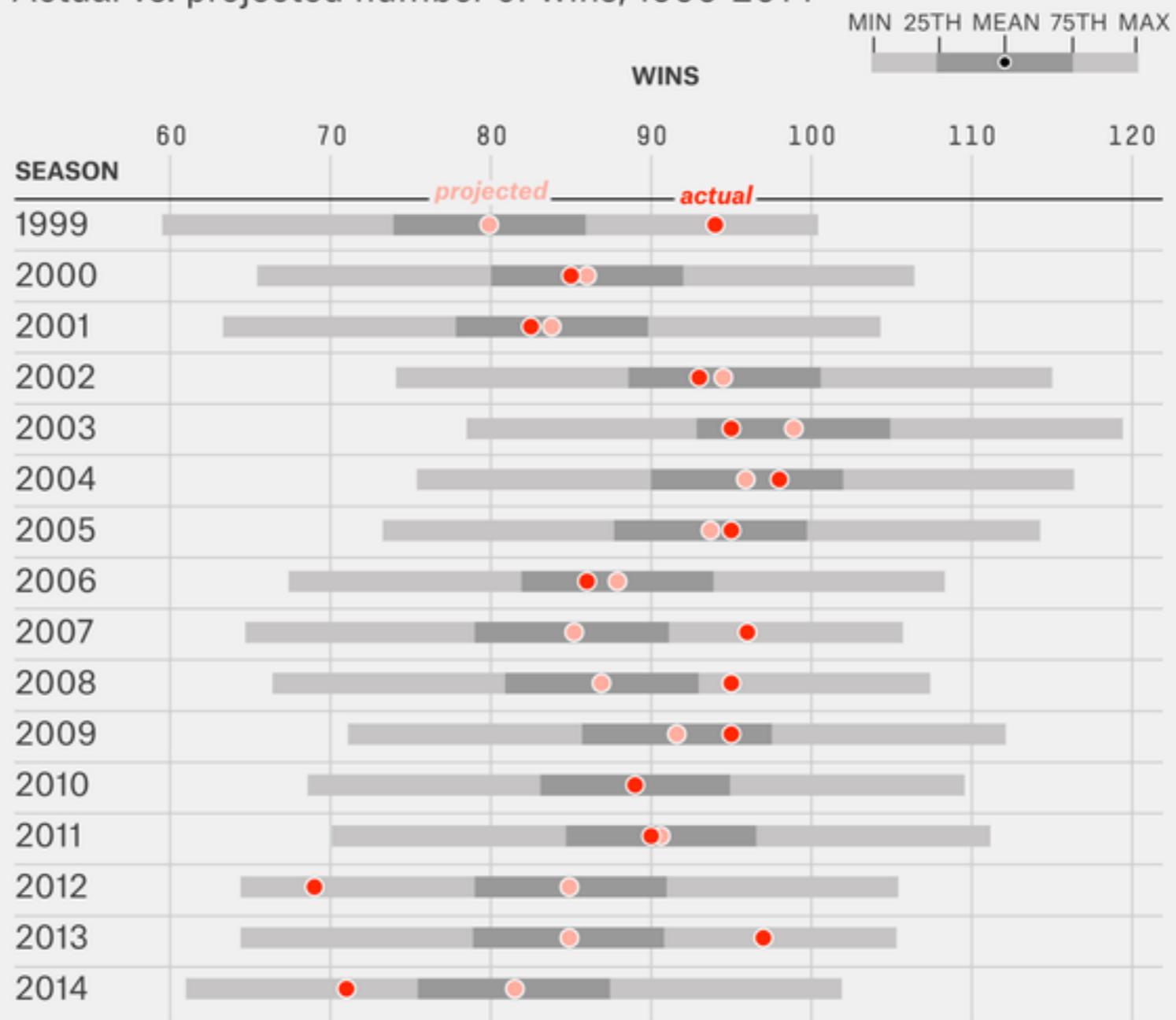


WAPO.ST/WONKBLOG

Source: U.S. Dept. of Education College Scorecard

Red Sox Roller Coaster

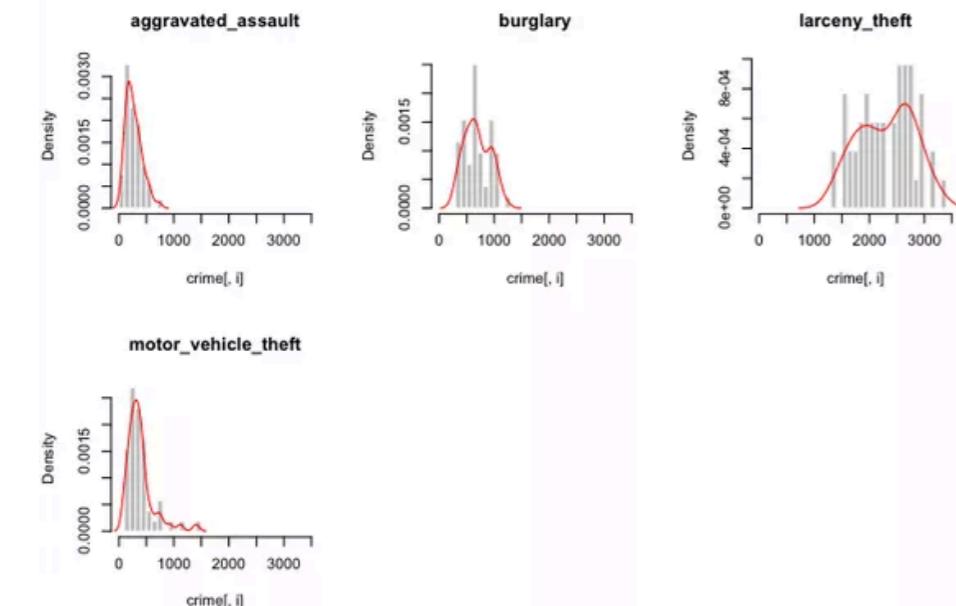
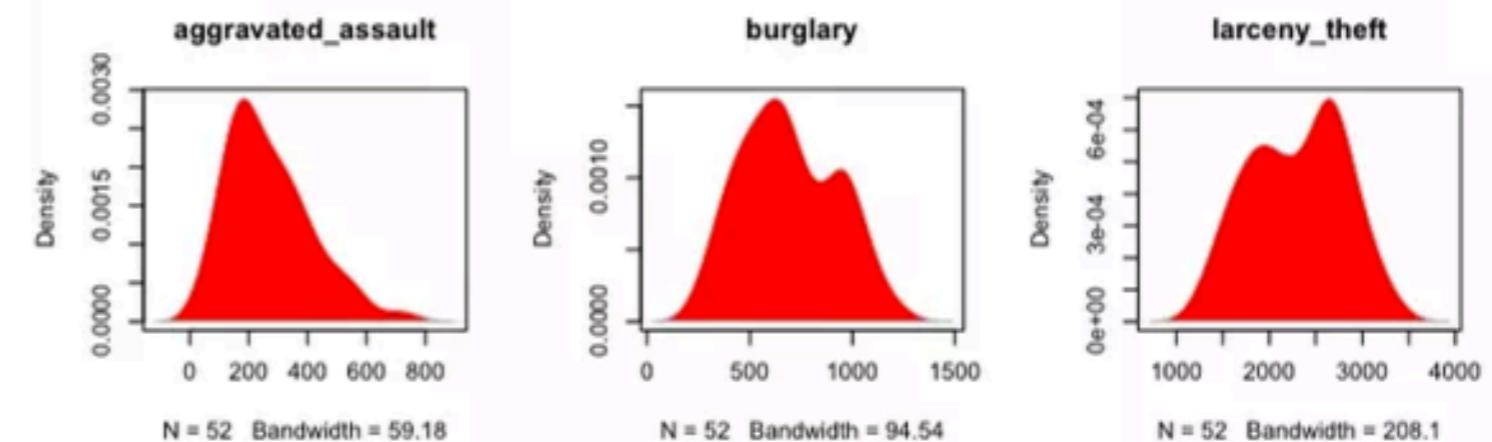
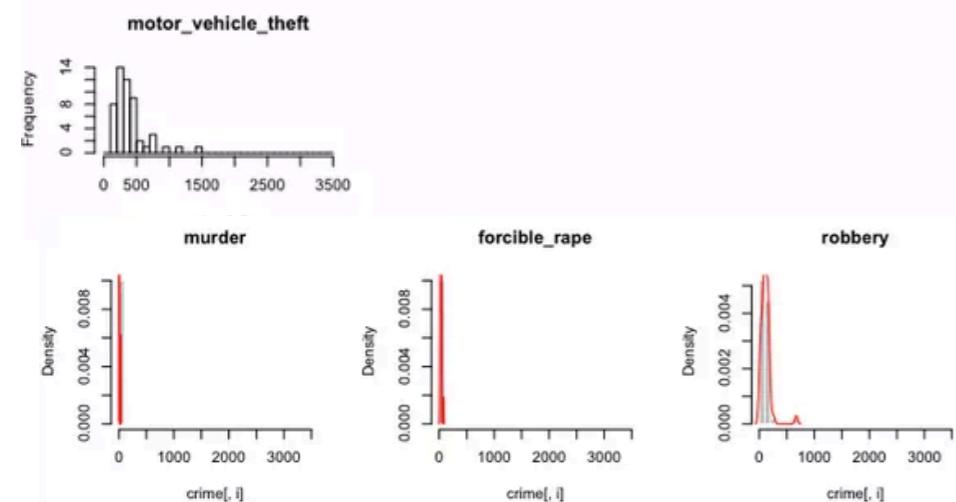
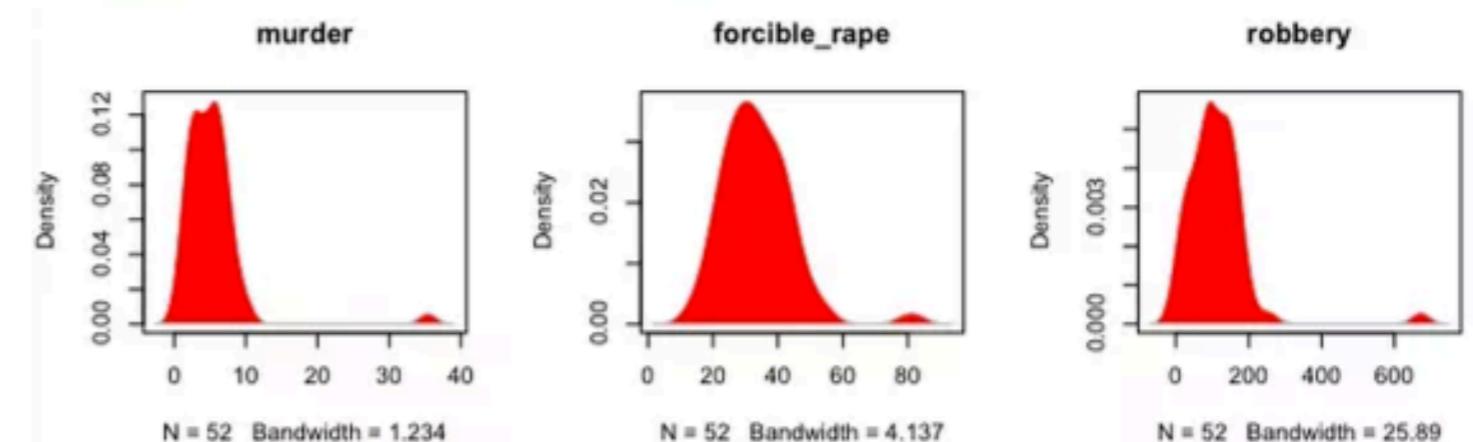
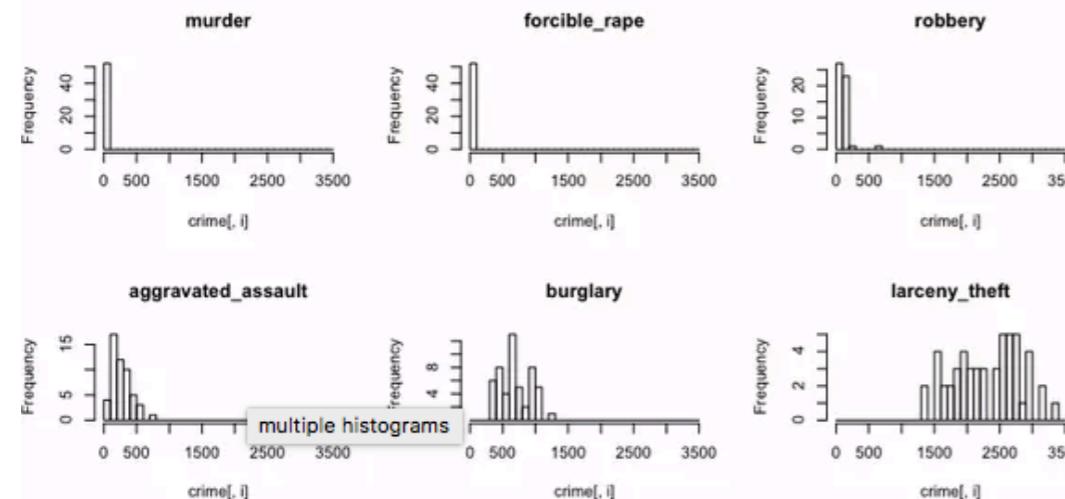
Actual vs. projected number of wins, 1999-2014



FIVETHIRTYEIGHT

SOURCE: BASEBALL-REFERENCE.COM, LAHMAN DB

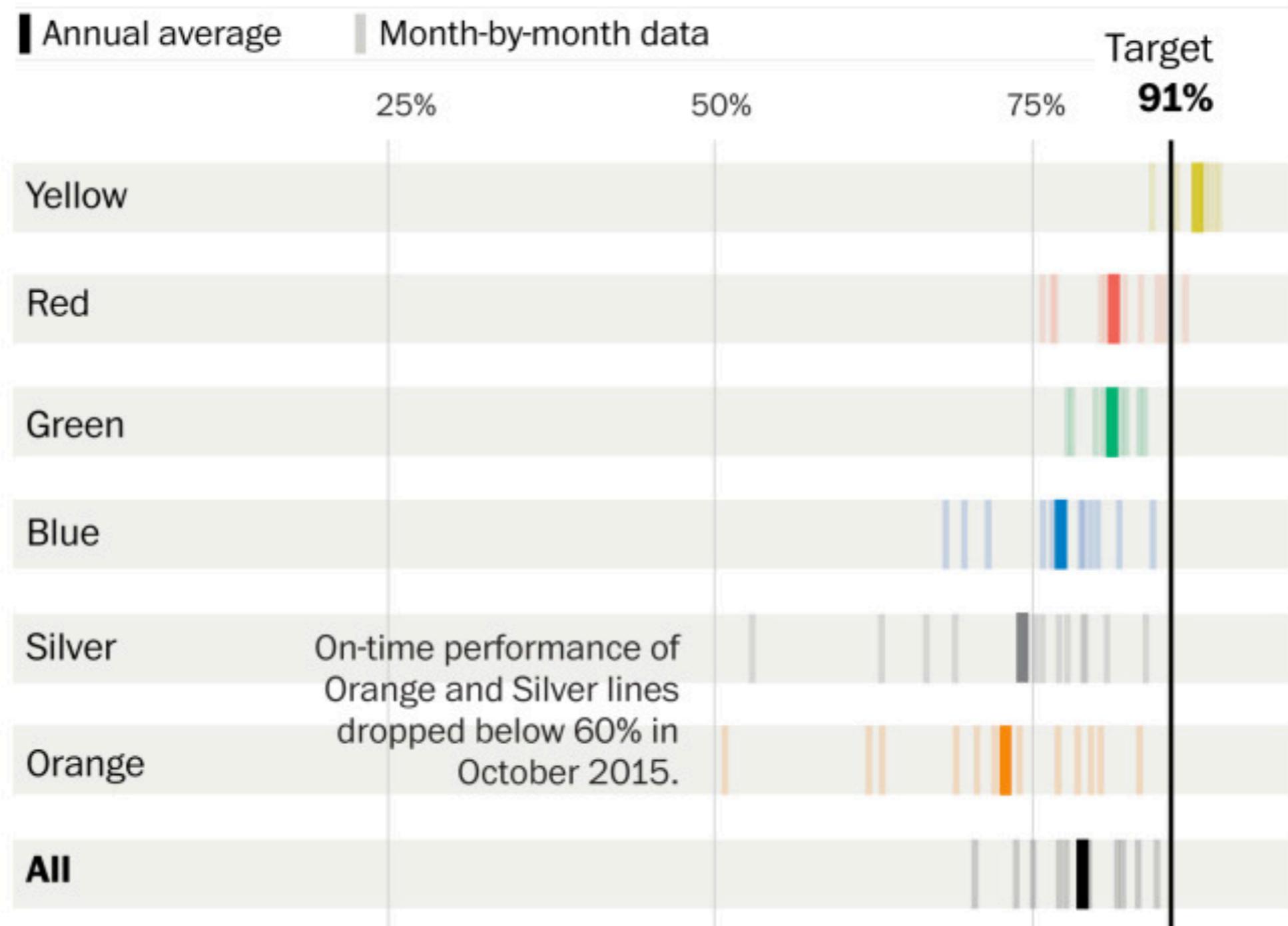
GRÁFICO DE DENSIDADE



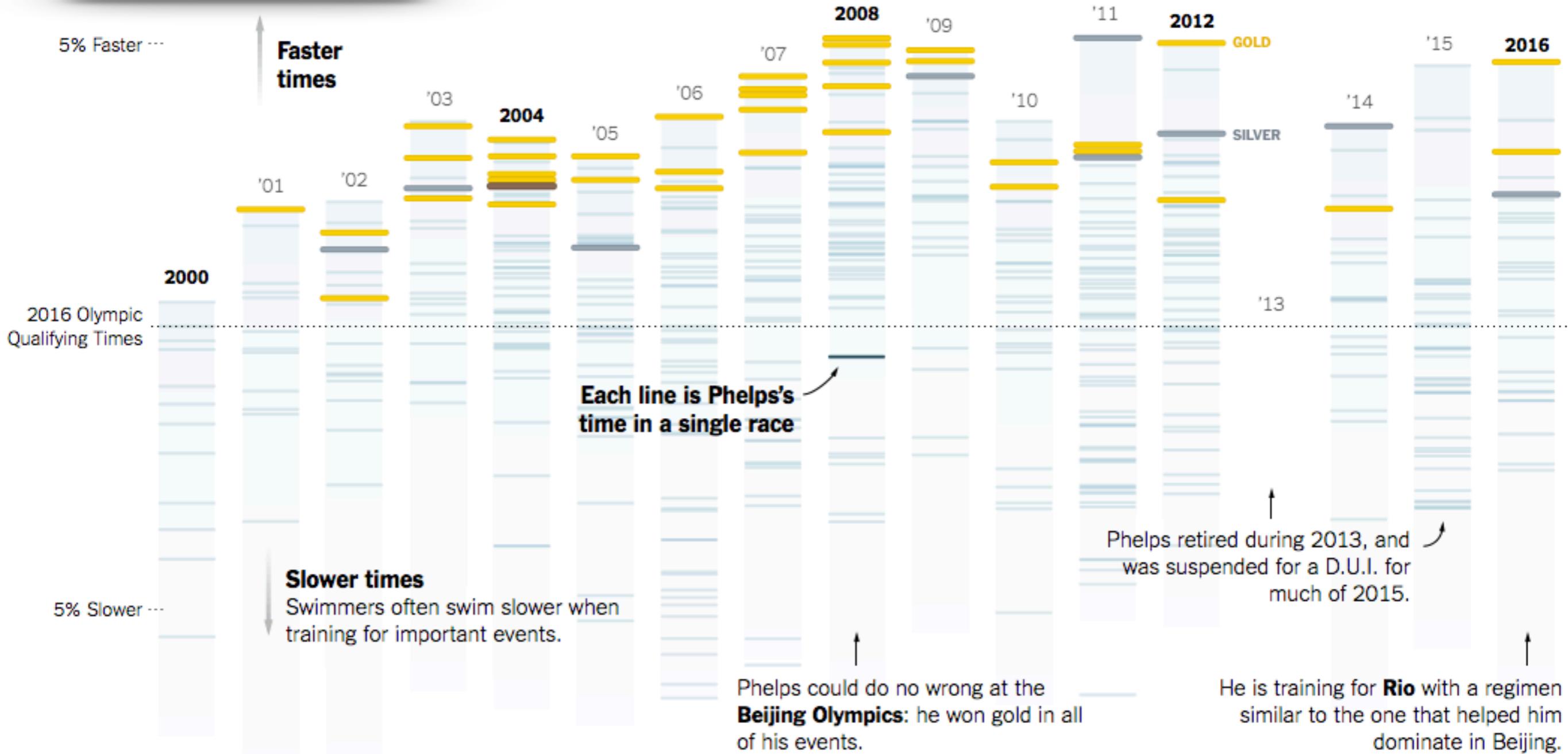
RUG PLOT OU BAR CODE

The system has performance issues

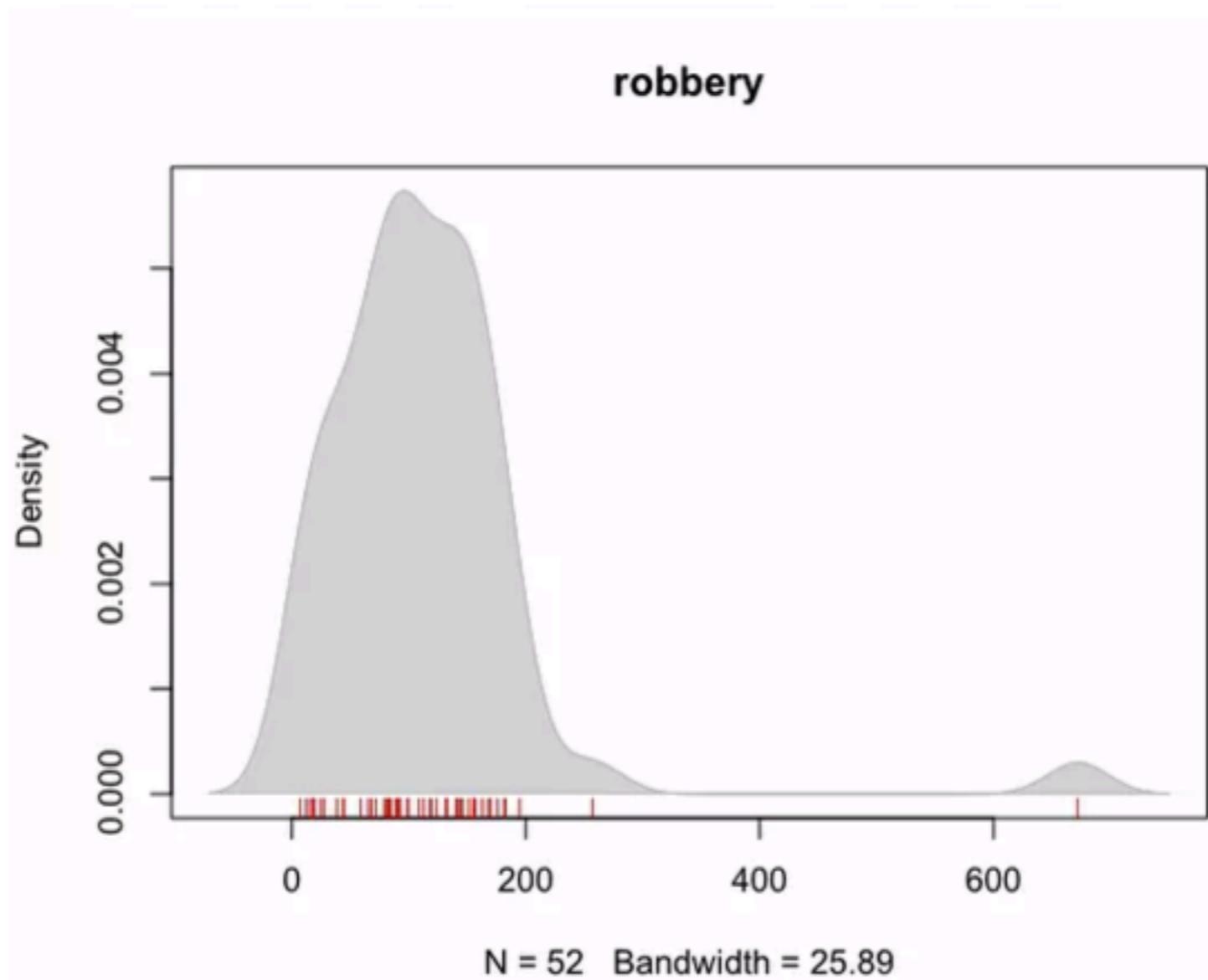
On-time performance by line in 2015



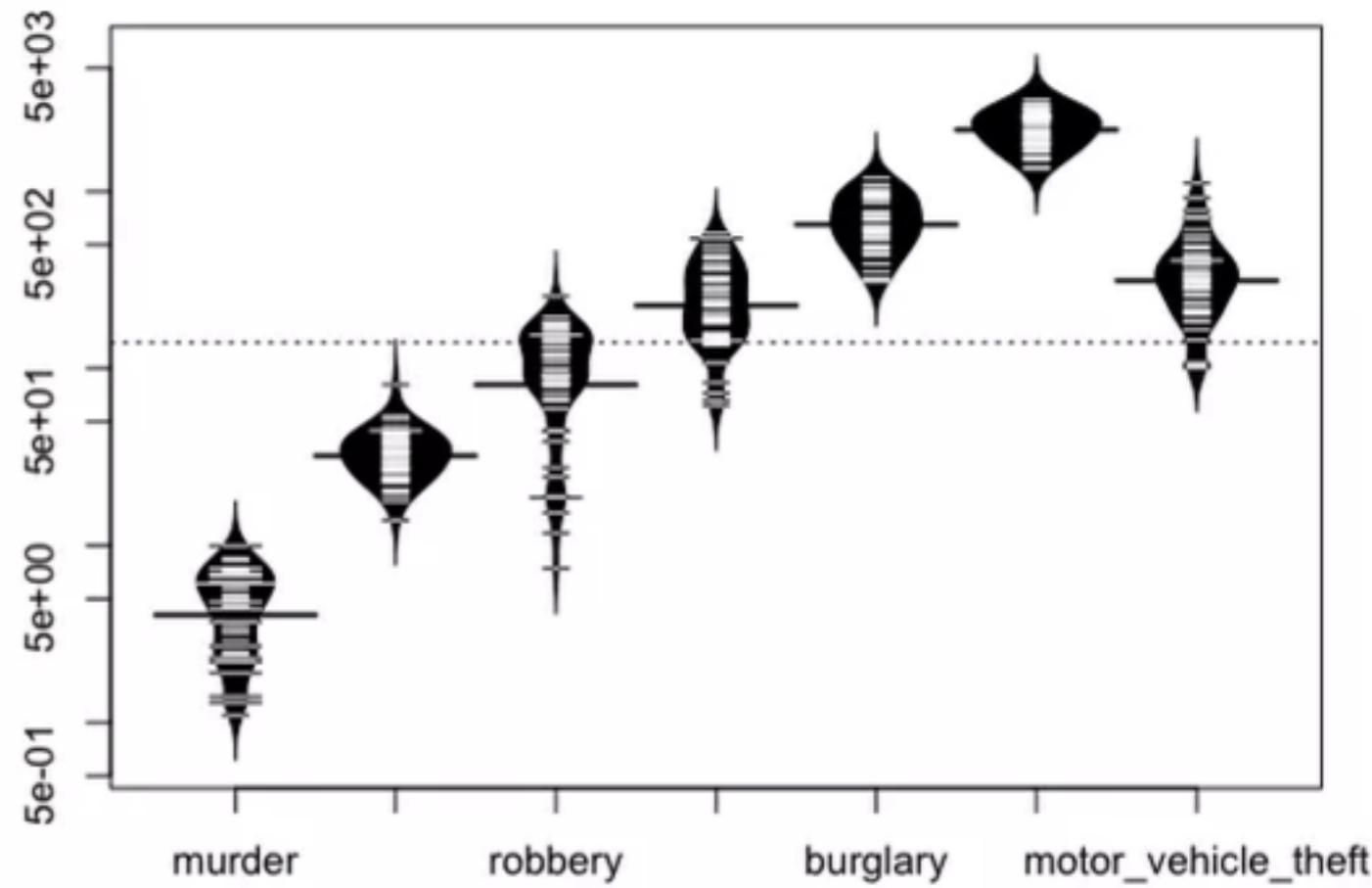
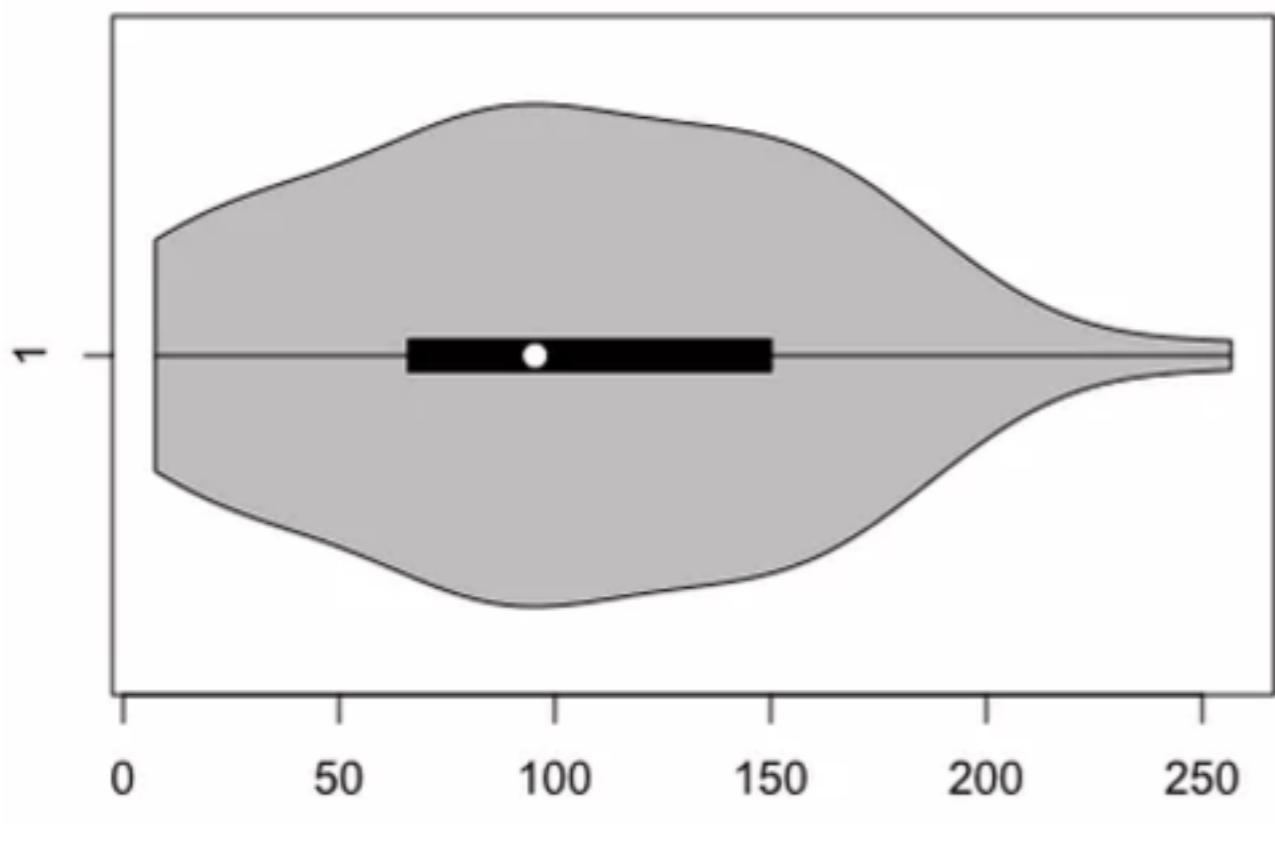
RUG PLOT OU BAR CODE



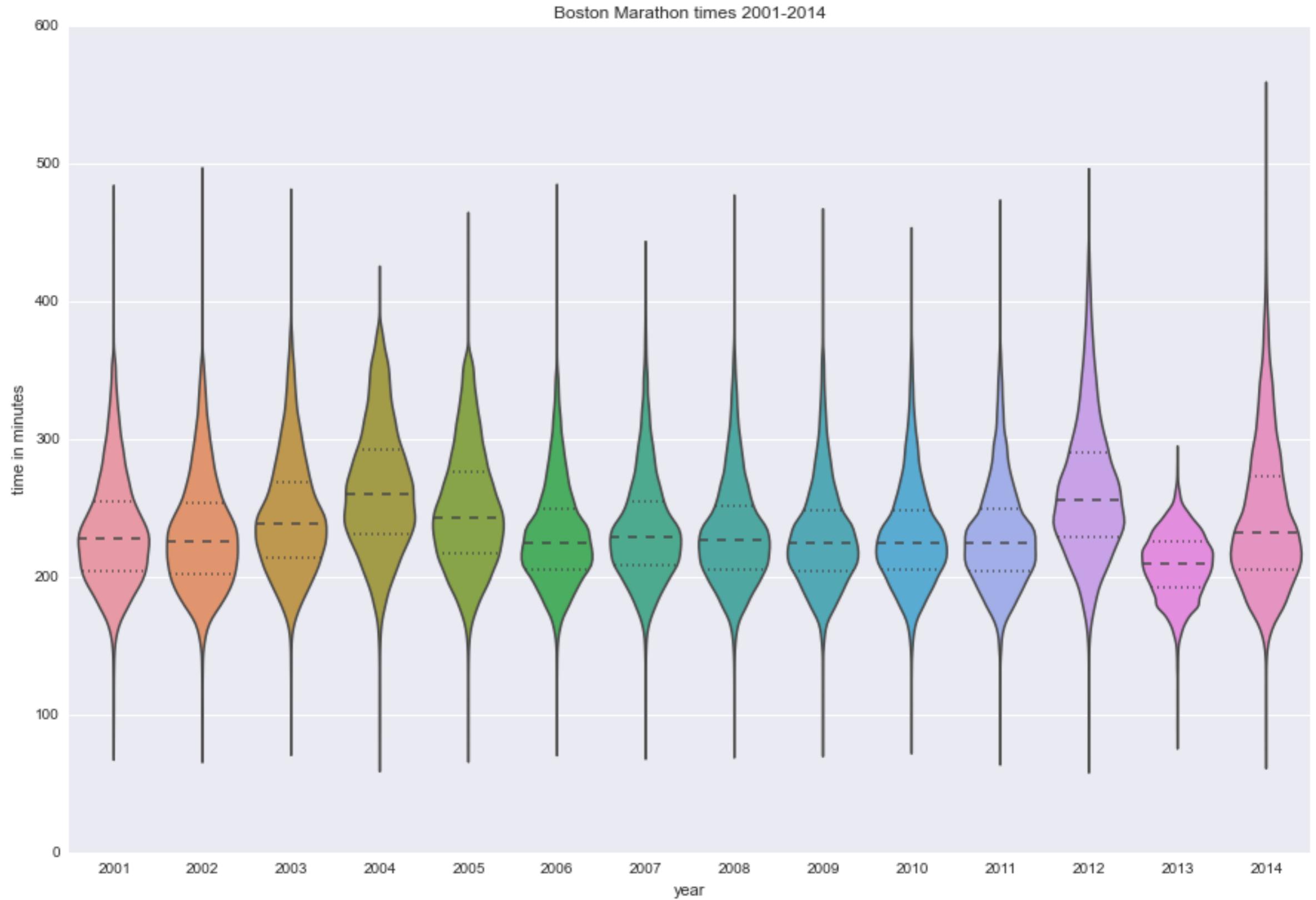
RUG PLOT OU BAR CODE



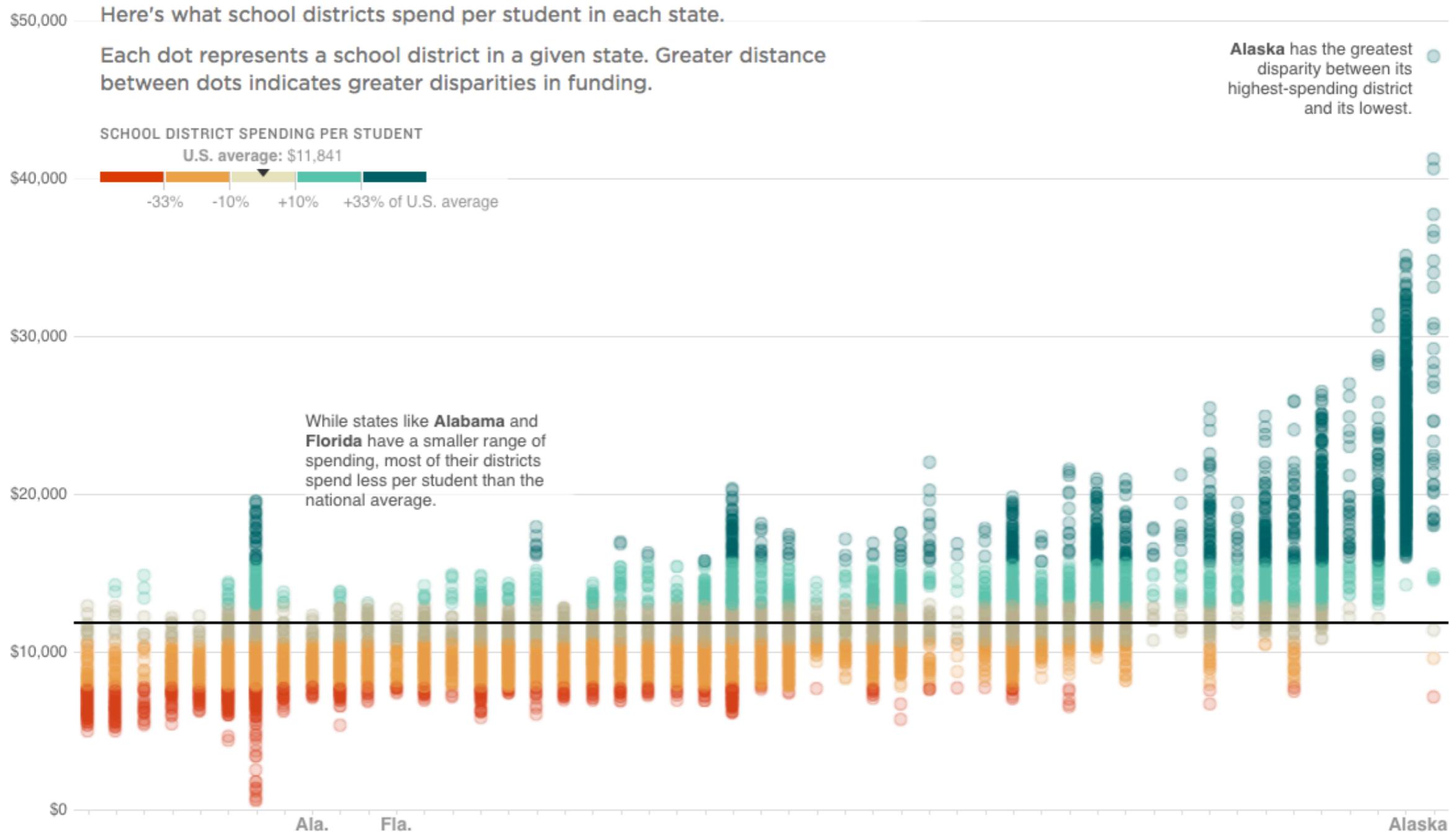
VIOLIN / BEAN PLOT



VIOLIN / BEAN PLOT



STRIP PLOT



Notes

This [Education Week](#) analysis of federal and state data excludes extreme outliers as well as districts with fewer than 200 students. Hawaii and Washington, D.C., are excluded because each has only one school district.

Source: [Education Week](#) analysis of federal and state data.

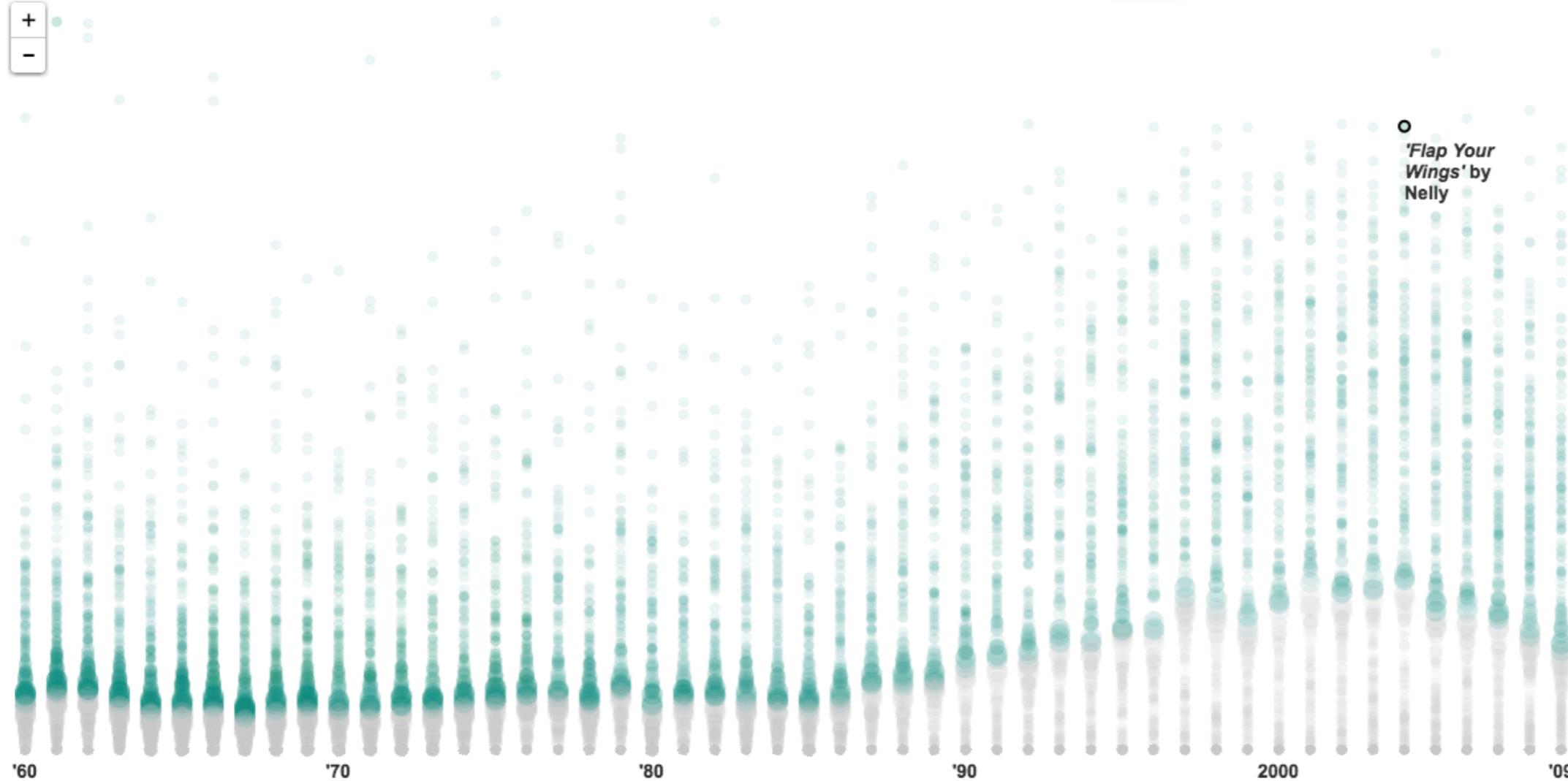
Credit: Katie Park, Alyson Hurt and Lisa Charlotte Rost/NPR

STRIP PLOT

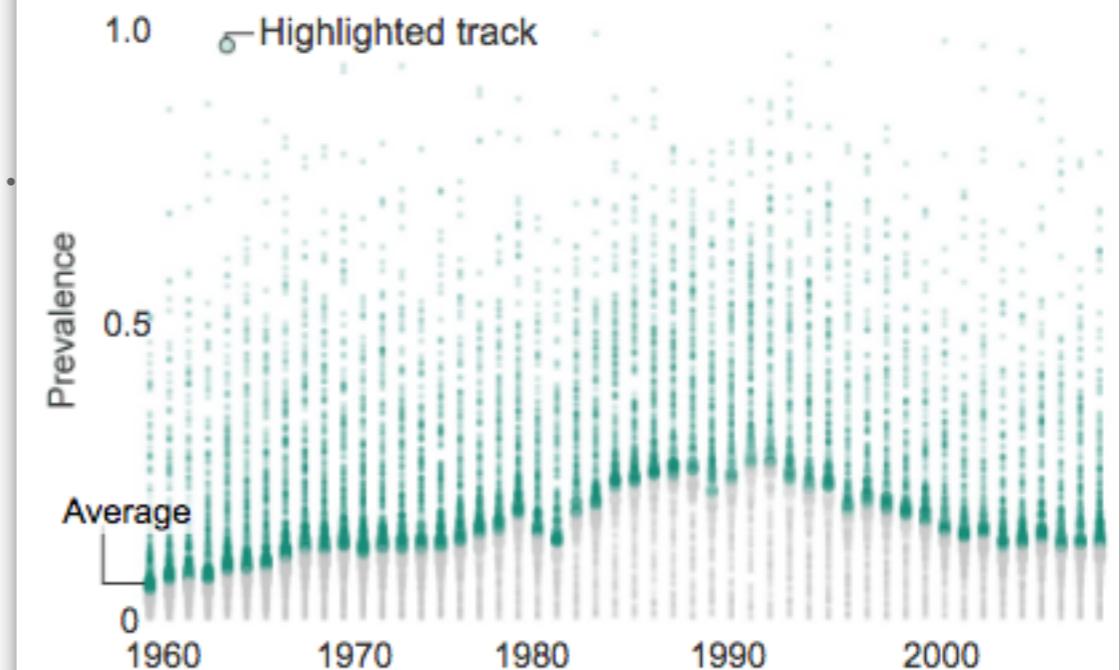
Hip-hop revolution ignites

Ramps up with the rise of rap and hip-hop in the late '80s and into the '90s, with artists like [Busta Rhymes](#), [Ludacris](#) and [Snoop Dogg](#).

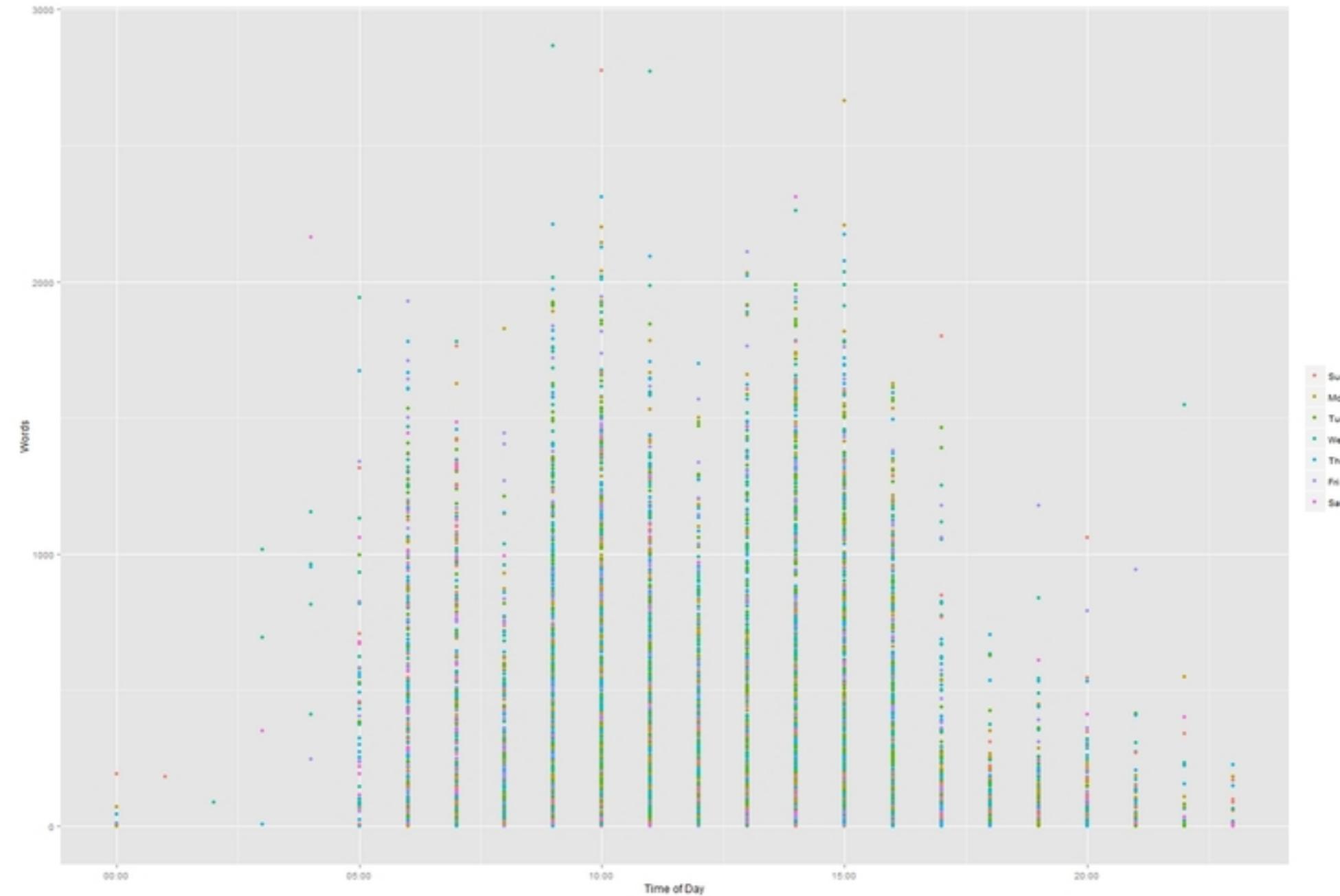
Elements: Energetic, speech, bright



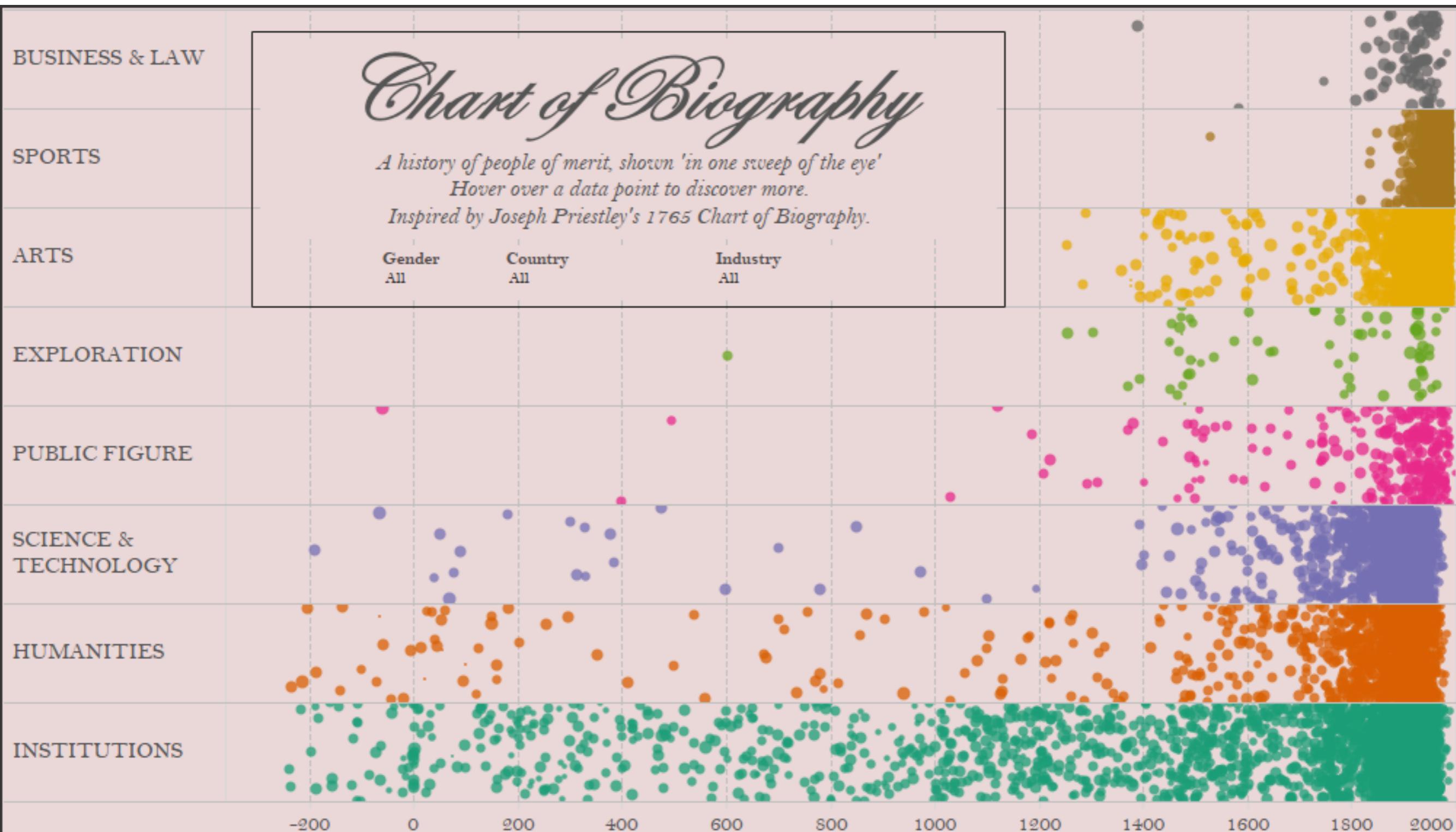
Reading this chart:



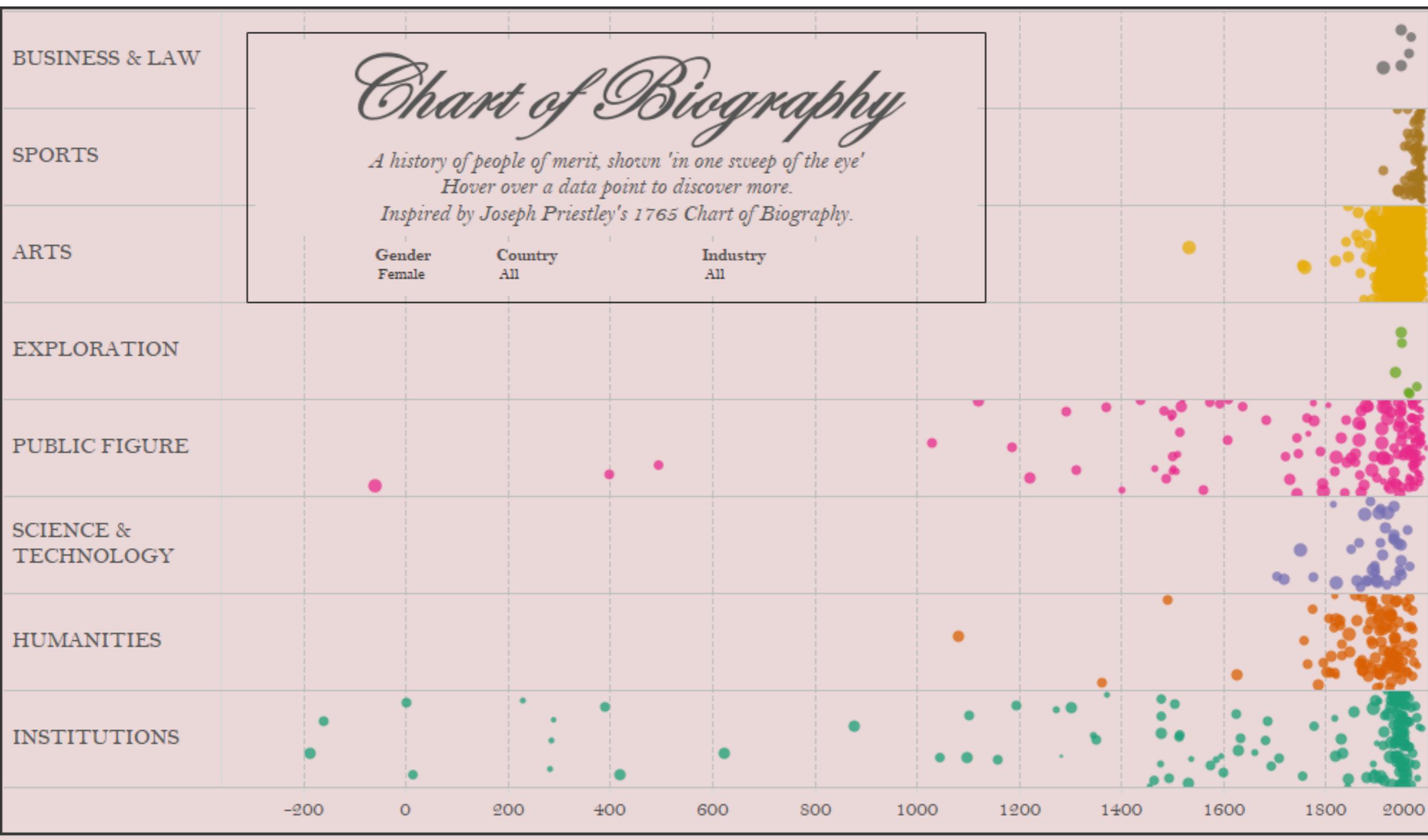
STRIP PLOT



JITTERED STRIP PLOT

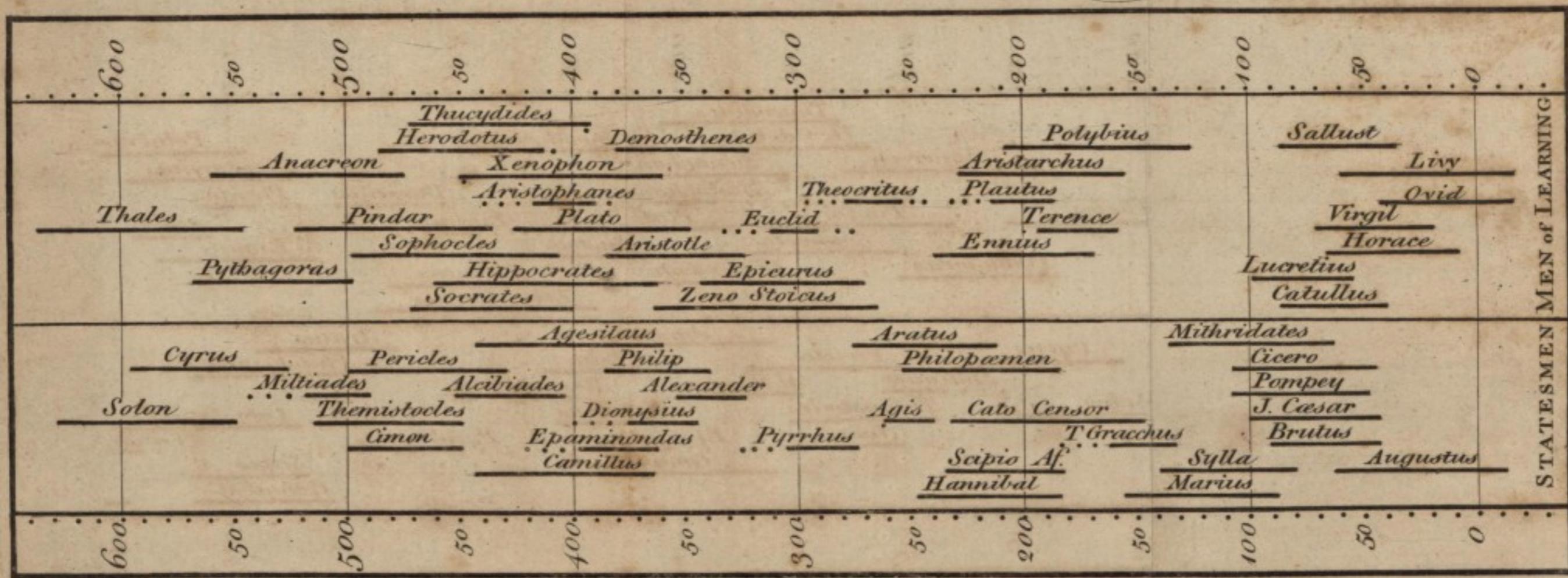


JITTERED STRIP PLOT

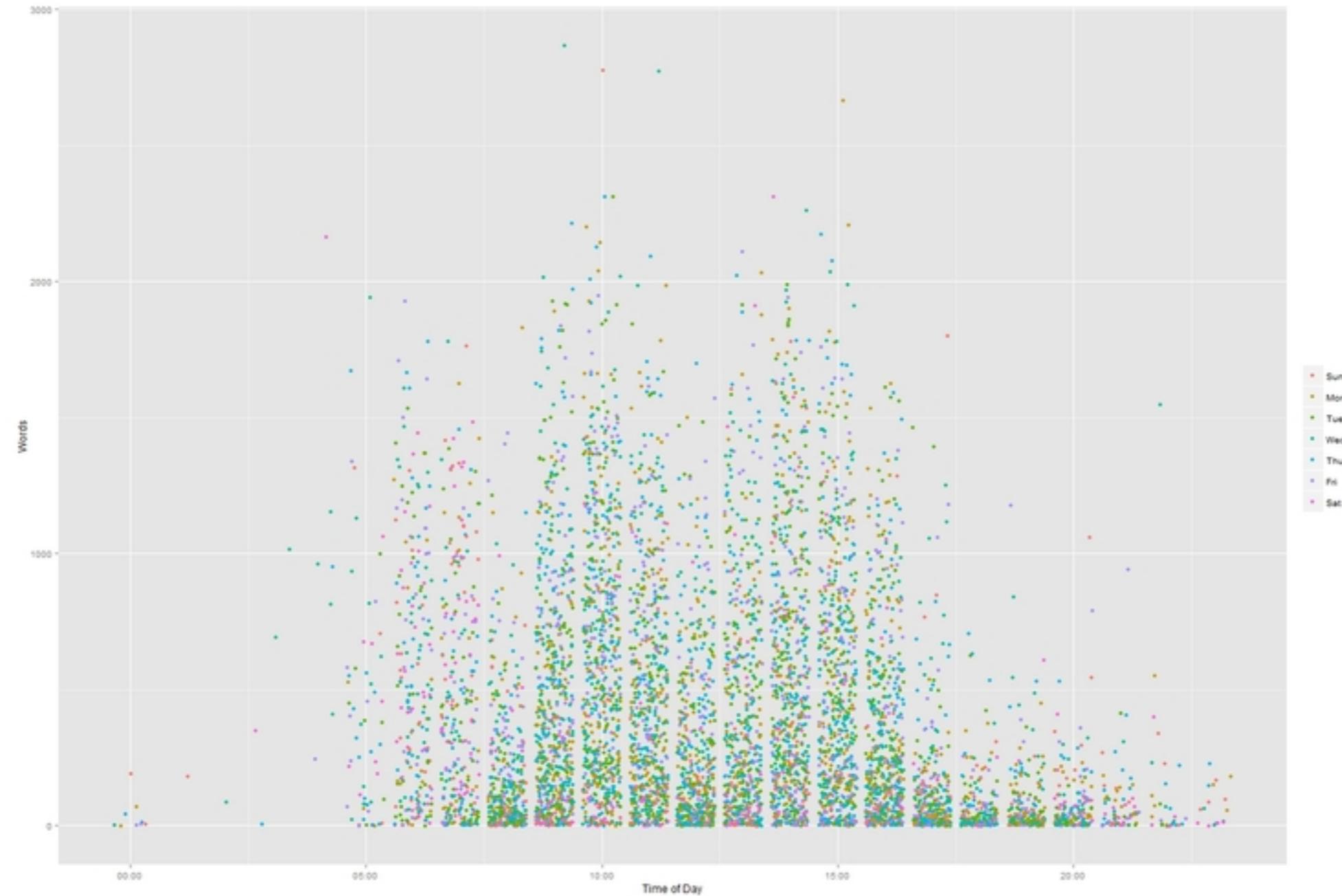


JITTERED STRIP PLOT

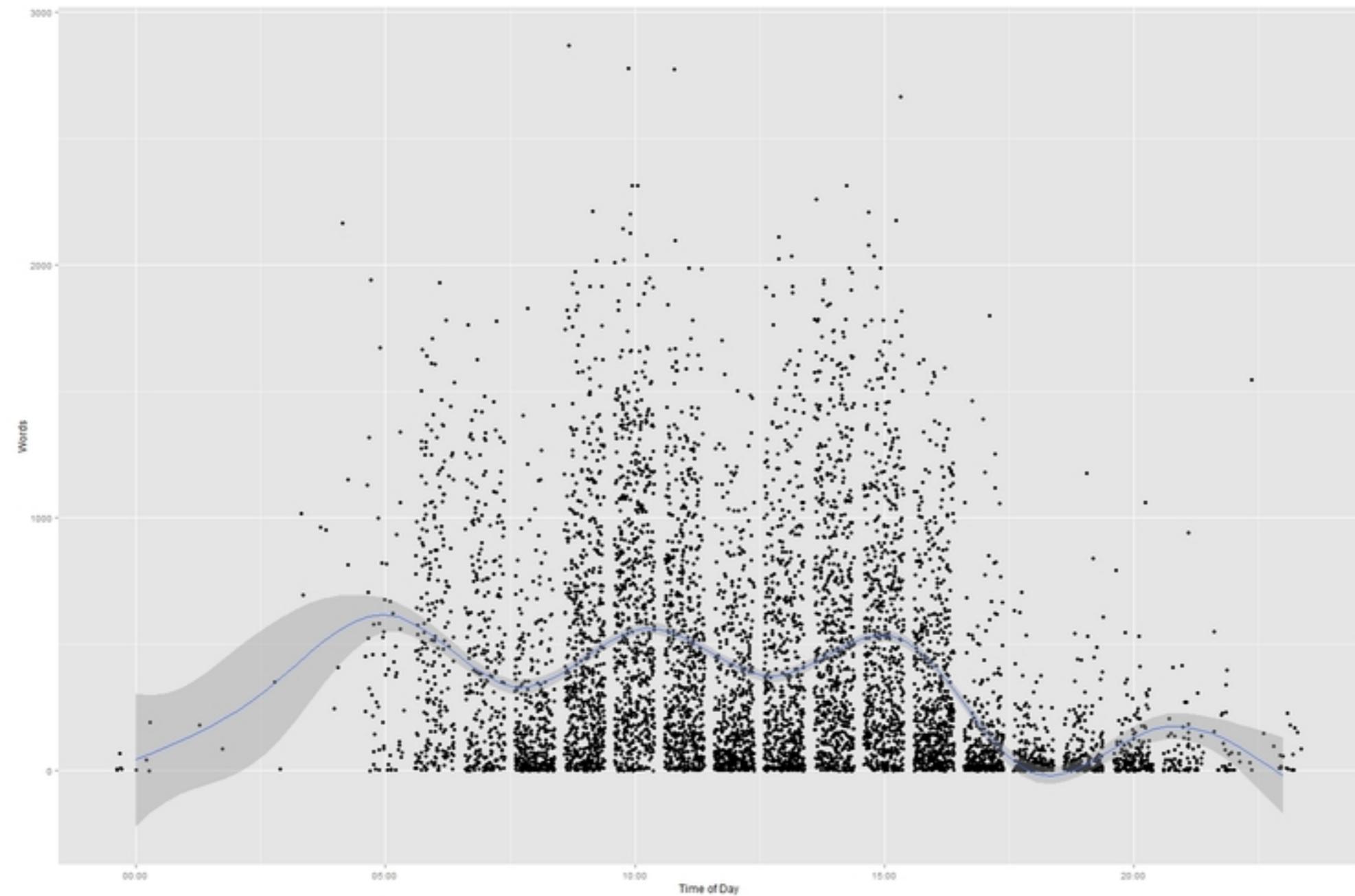
A Specimen of a Chart of Biography.



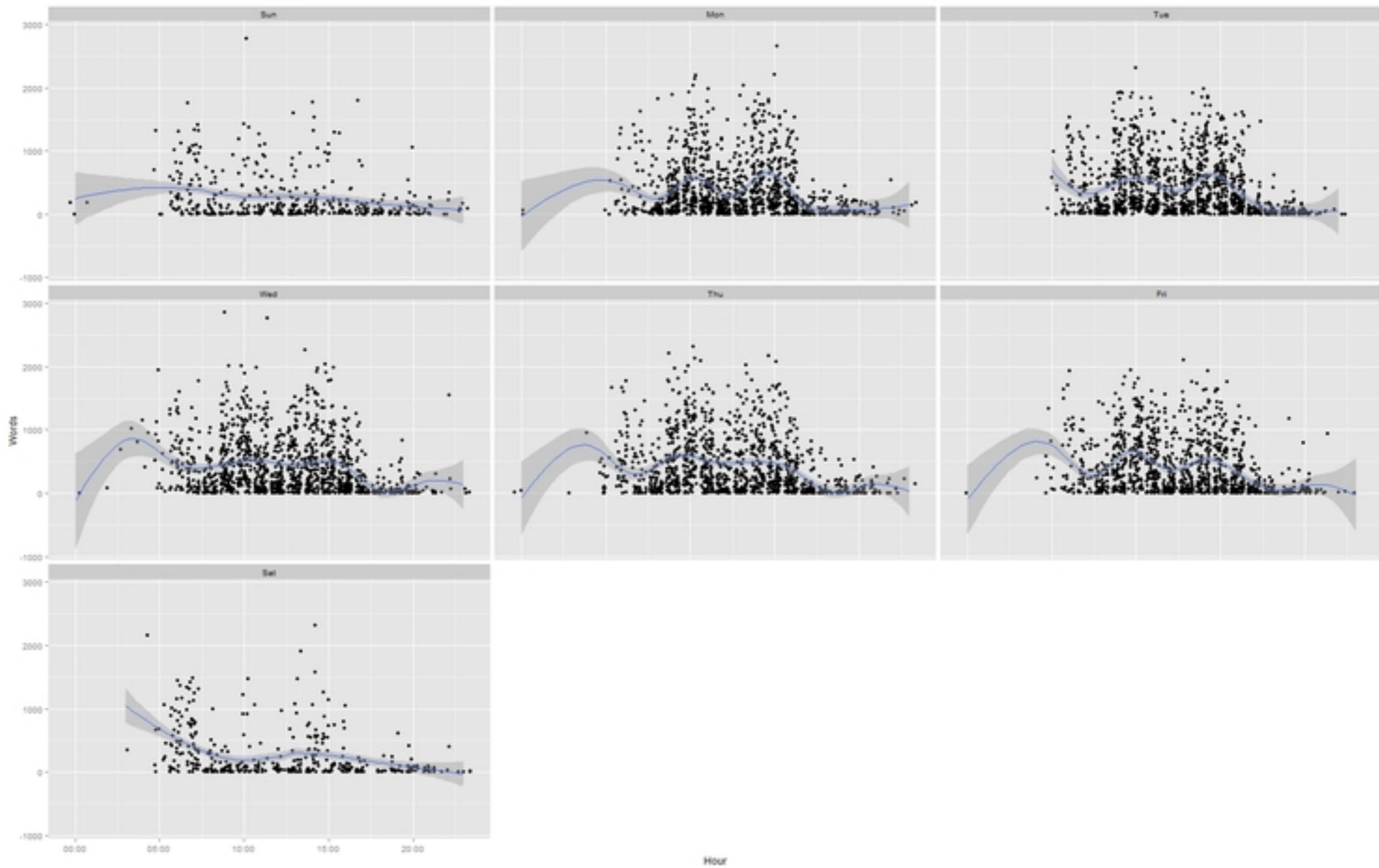
JITTERED STRIP PLOT



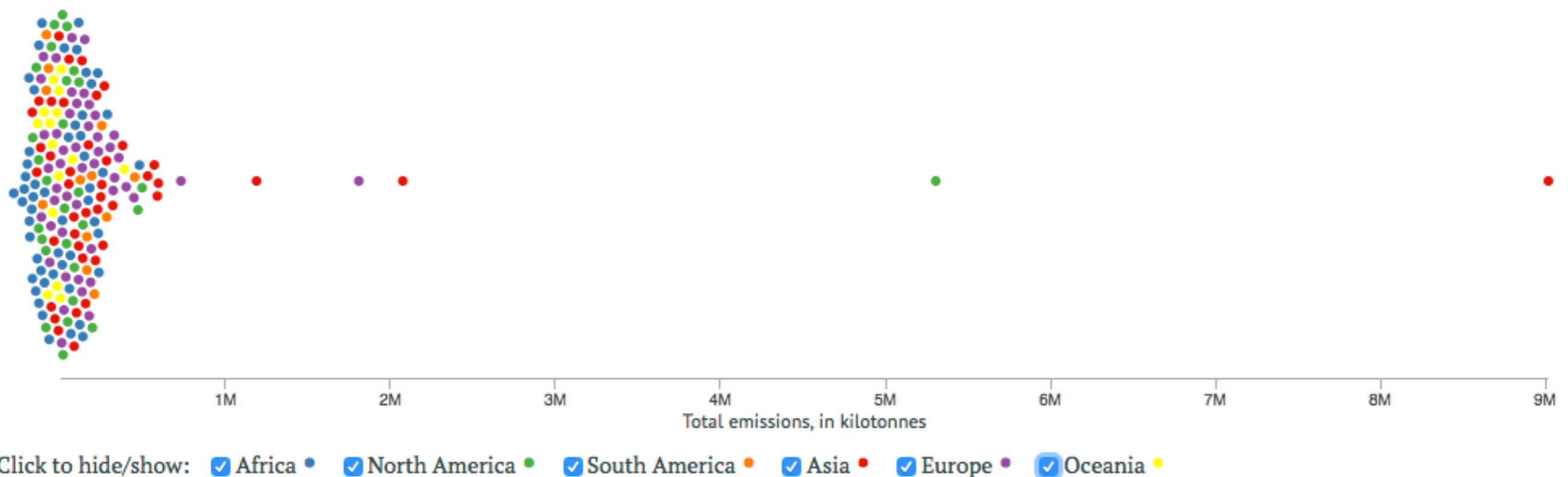
JITTERED STRIP PLOT



JITTERED STRIP PLOT



BEE SWARM PLOT

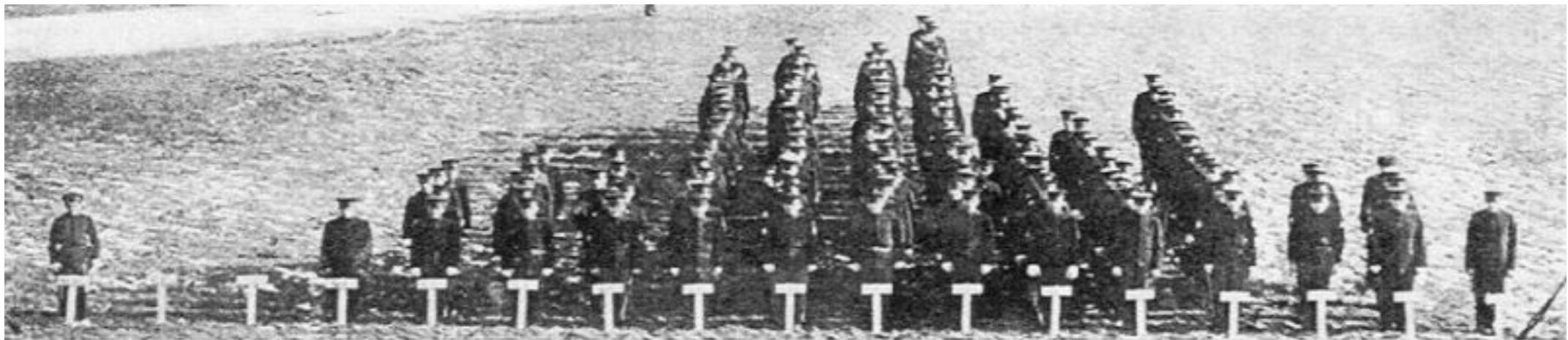


BEE SWARM PLOT

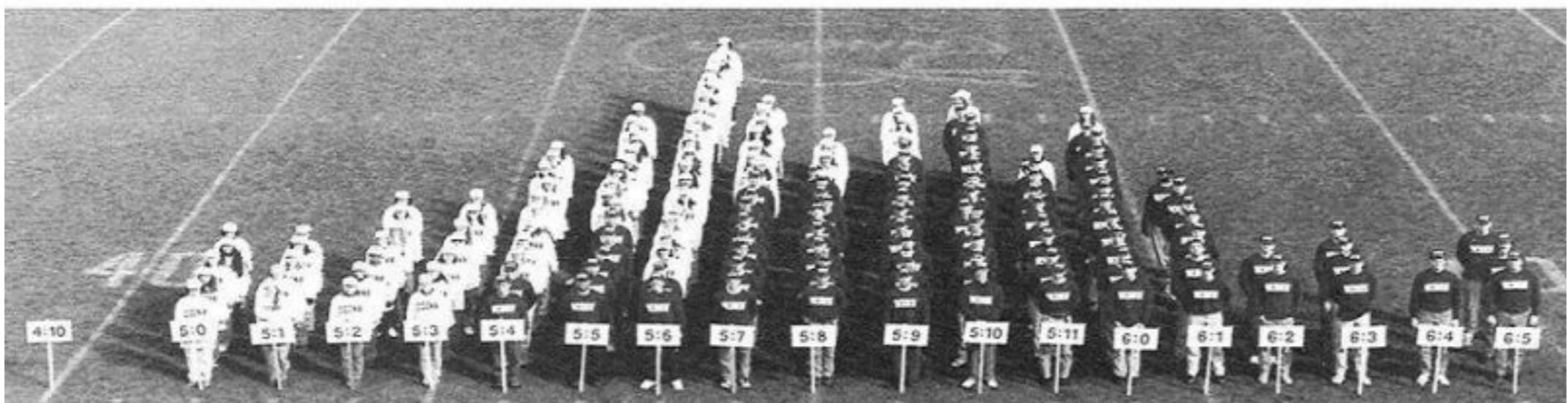
SELECT YEAR 1960 1980 2000 2014



UNIT PLOT

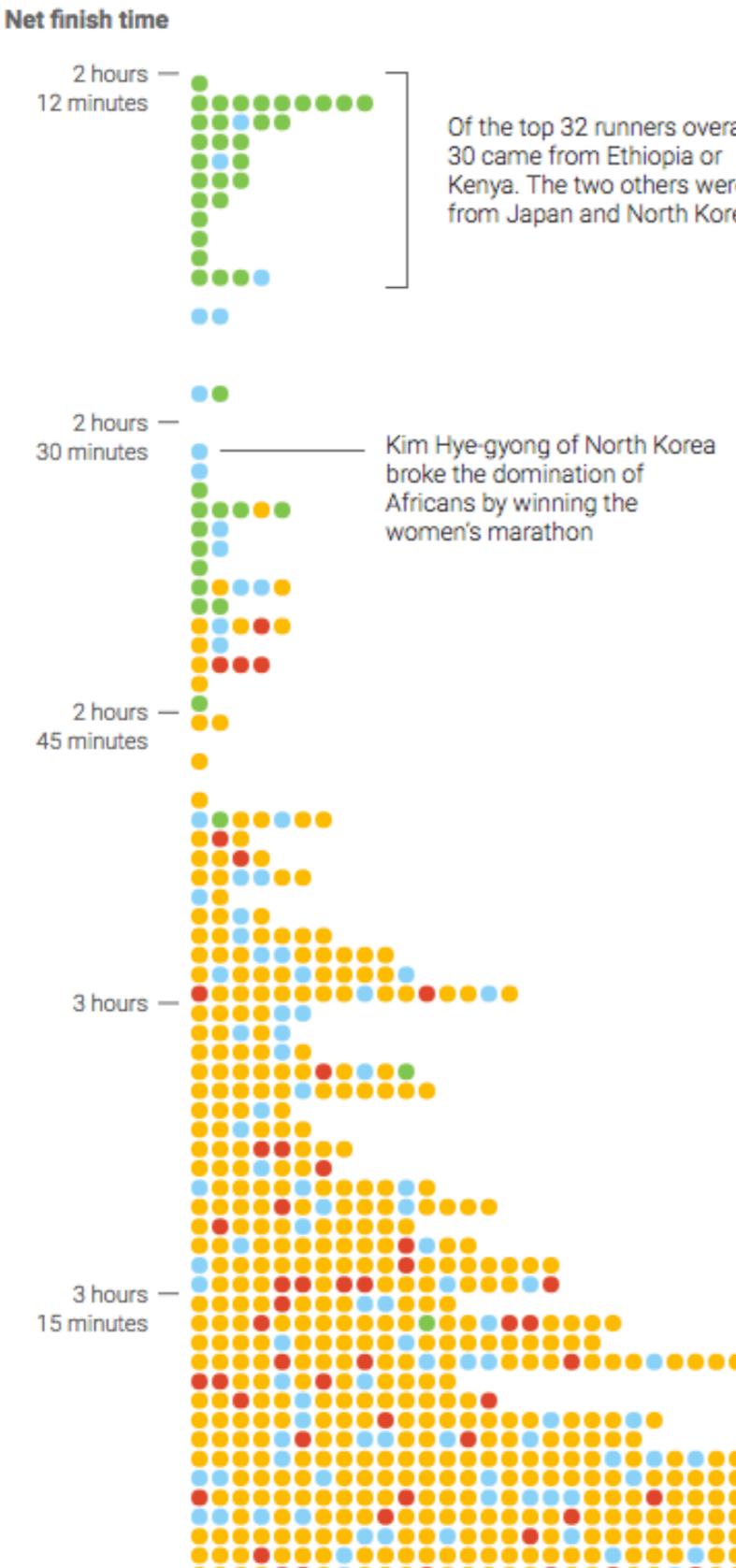


4:10 4:11 5:0 5:1 5:2 5:3 5:4 5:5 5:6 5:7 5:8 5:9 5:10 5:11 6:0 6:1 6:2



4:10 5:0 5:1 5:2 5:3 5:4 5:5 5:6 5:7 5:8 5:9 5:10 5:11 6:0 6:1 6:2 6:3 6:4 6:5

UNIT PLOT

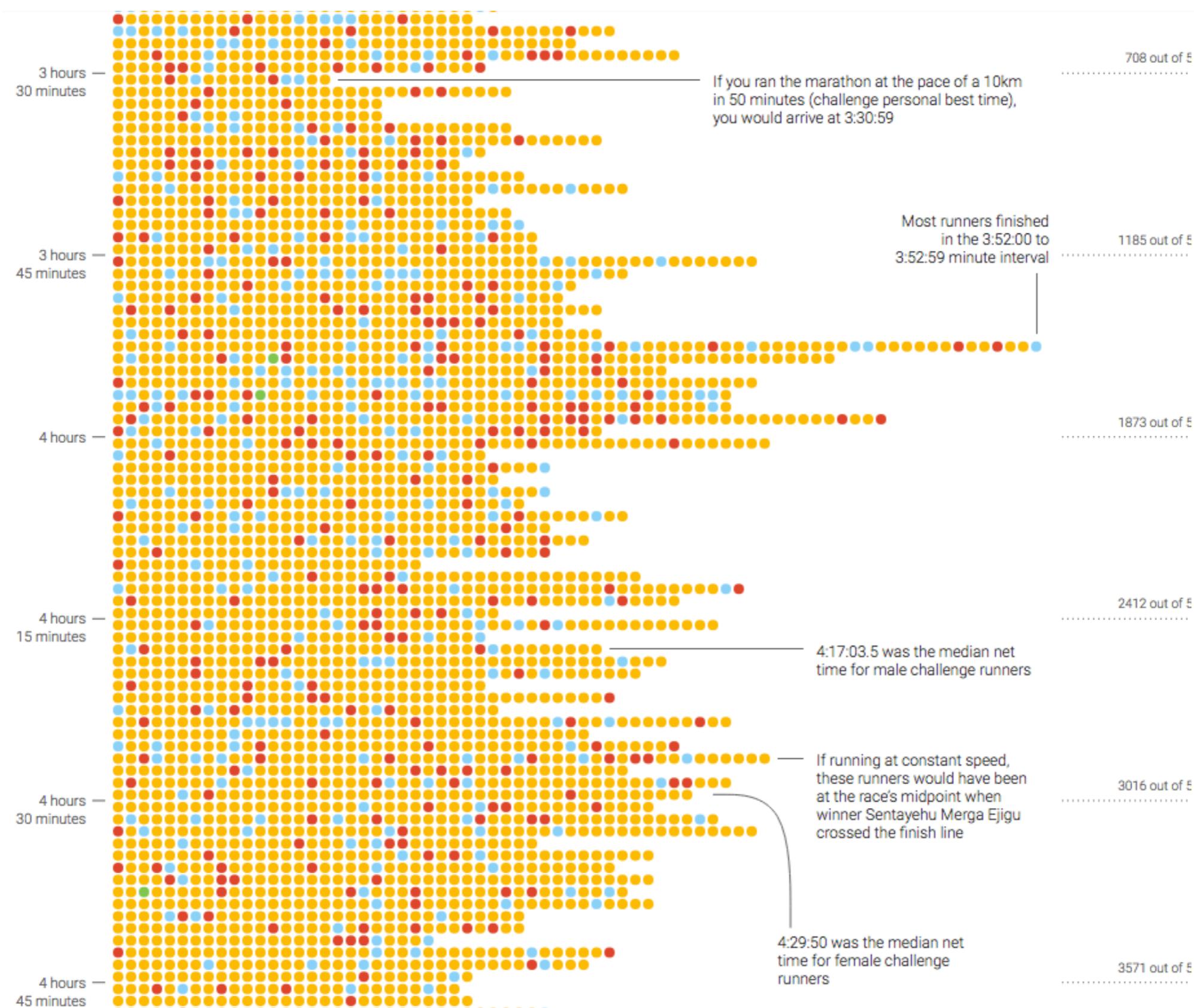


In the long run (2015 revision)

By Cedric Sam

This year's Standard Chartered Hong Kong Marathon was won on Sunday morning by Ethiopia's Sentayehu Merga Ejigu in 2:12:59 (net time), almost 2 minutes better than the best score in 2014. In the women's competition, Kim Hye-gyong of North Korea broke the domination of Africans by winning in 2:31:46. We reprise our 2014 graphic and again trace the milestones of the marathon challenge, the toughest among the event's races, from the firing gun to the last second. (Published On 2015-01-26)

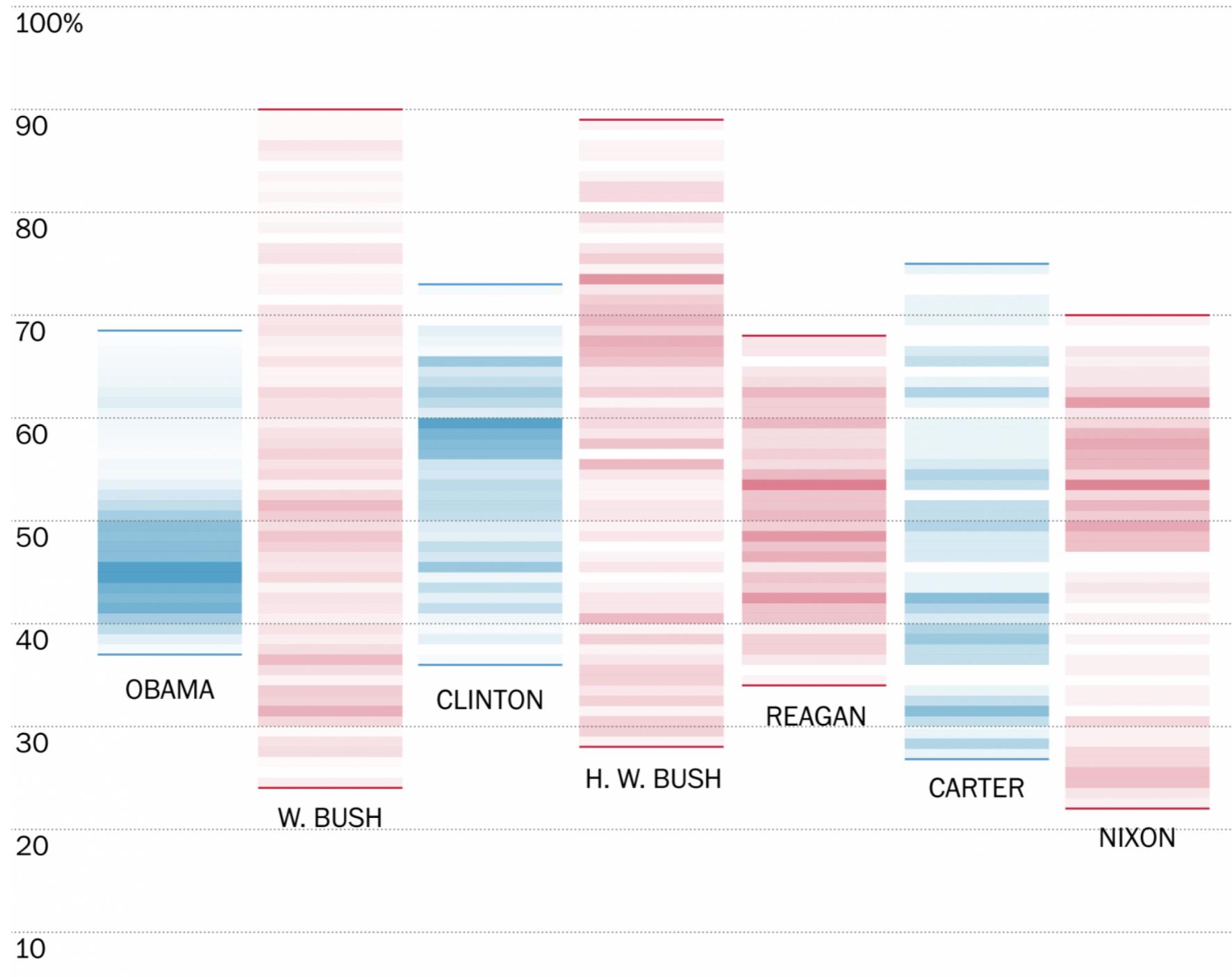
UNIT PLOT



MAPA DE CALOR

Presidential approval frequency

According to polling from Gallup.



MAPA DE CALOR

Utilization of Squares by Chess Masters

Alexander Alekhine

Playing as White				Playing as Black				
8	0.12	0.12	0.16	0.19	0.13	0.11	0.08	0.07
7	0.26	0.27	0.27	0.35	0.38	0.28	0.17	0.18
6	0.27	0.32	0.48	0.60	0.50	0.58	0.38	0.29
5	0.36	0.77	0.72	1.36	1.30	0.77	0.96	0.42
4	0.72	0.58	1.52	2.04	1.80	1.24	0.67	0.64
3	0.48	0.81	1.71	1.21	1.32	1.95	0.84	0.52
2	0.10	0.21	0.62	1.01	1.02	0.43	0.40	0.18
1	0.16	0.34	0.74	0.94	0.69	1.16	0.95	0.22
a	b	c	d	e	f	g	h	

Viswanathan Anand

Playing as White				Playing as Black				
8	0.15	0.11	0.14	0.22	0.14	0.10	0.07	0.07
7	0.13	0.55	0.58	1.29	1.47	0.56	0.54	0.22
6	0.65	0.90	1.77	1.26	1.48	2.15	0.96	0.63
5	0.74	0.76	1.46	1.86	1.56	1.17	0.73	0.65
4	0.40	0.88	0.73	1.07	1.04	0.65	0.71	0.39
3	0.23	0.36	0.63	0.46	0.43	0.43	0.28	0.26
2	0.22	0.32	0.30	0.33	0.26	0.24	0.21	0.16
1	0.12	0.12	0.16	0.22	0.18	0.15	0.08	0.06
a	b	c	d	e	f	g	h	

José Raúl Capablanca

Playing as White				Playing as Black				
8	0.09	0.09	0.15	0.25	0.20	0.12	0.03	0.03
7	0.21	0.27	0.25	0.32	0.47	0.19	0.19	0.11
6	0.23	0.30	0.54	0.46	0.48	0.56	0.34	0.22
5	0.32	0.80	0.84	1.19	1.12	0.68	1.05	0.38
4	0.76	0.70	1.57	1.82	1.58	1.05	0.76	0.69
3	0.51	0.89	1.70	1.27	2.03	0.89	0.60	
2	0.09	0.27	0.62	1.04	1.05	0.44	0.59	0.24
1	0.13	0.34	0.82	0.79	0.76	1.24	1.02	0.13
a	b	c	d	e	f	g	h	

Magnus Carlsen

Playing as White				Playing as Black				
8	0.14	0.16	0.18	0.27	0.19	0.15	0.08	0.10
7	0.24	0.33	0.30	0.39	0.30	0.28	0.21	0.17
6	0.34	0.38	0.60	0.61	0.51	0.55	0.34	0.28
5	0.47	0.89	0.80	1.40	1.28	0.77	0.88	0.44
4	0.88	0.75	1.64	2.25	1.82	1.24	0.83	0.81
3	0.55	1.00	1.81	1.21	1.52	2.20	1.09	0.65
2	0.15	0.37	0.77	1.31	1.27	0.63	0.75	0.32
1	0.19	0.48	0.83	1.09	0.91	1.46	1.09	0.39
a	b	c	d	e	f	g	h	

Bobby Fischer

Playing as White				Playing as Black				
8	0.12	0.11	0.14	0.24	0.15	0.14	0.06	0.09
7	0.11	0.47	0.68	1.52	1.14	0.55	1.26	0.26
6	0.80	0.91	1.35	1.28	1.32	2.25	1.18	0.65
5	0.35	0.96	0.66	1.06	1.15	0.75	0.86	0.45
4	0.87	0.65	1.27	2.04	1.79	1.18	0.86	0.66
3	0.55	1.10	1.73	0.95	1.24	2.05	0.92	0.65
2	0.09	0.22	0.46	1.20	1.06	0.42	0.50	0.25
1	0.18	0.37	0.60	0.98	0.82	1.36	1.04	0.24
a	b	c	d	e	f	g	h	

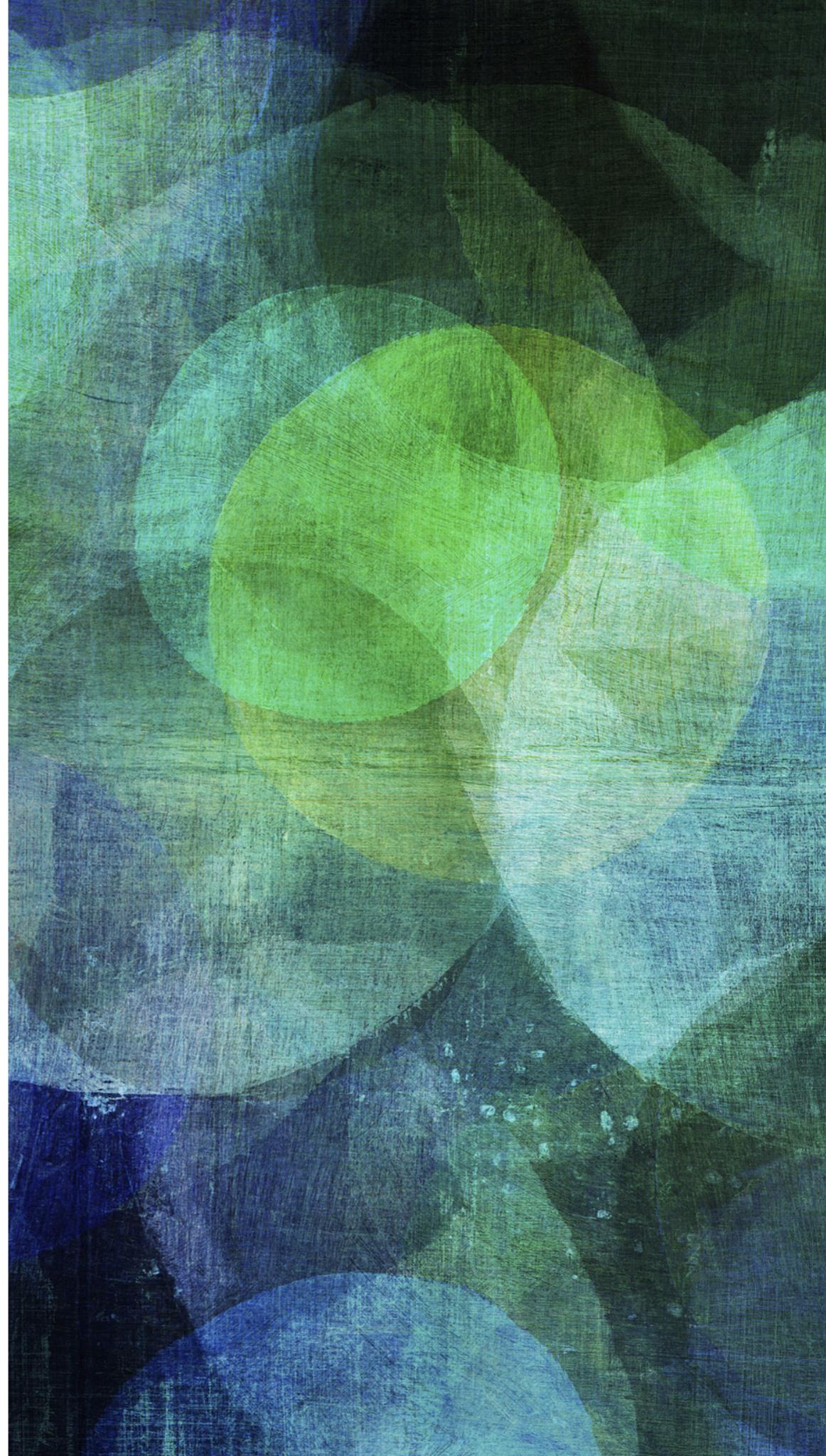
Anatoly Karpov

Playing as White				Playing as Black				
8	0.12	0.11	0.20	0.20	0.14	0.10	0.05	0.05
7	0.24	0.28	0.30	0.31	0.27	0.20	0.17	0.11
6	0.30	0.36	0.54	0.53	0.40	0.44	0.26	0.23
5	0.44	0.84	0.84	1.35	1.10	0.58	0.80	0.34
4	0.85	0.81	1.73	2.14	1.63	1.17	0.74	0.73
3	0.66	1.03	1.75	1.20	1.51	2.10	1.03	0.69
2	0.20	0.41	0.88	1.49	1.29	0.56	0.82	0.26
1	0.18	0.50	0.90	1.09	0.81	1.38	1.05	0.20
a	b	c	d	e	f	g	h	

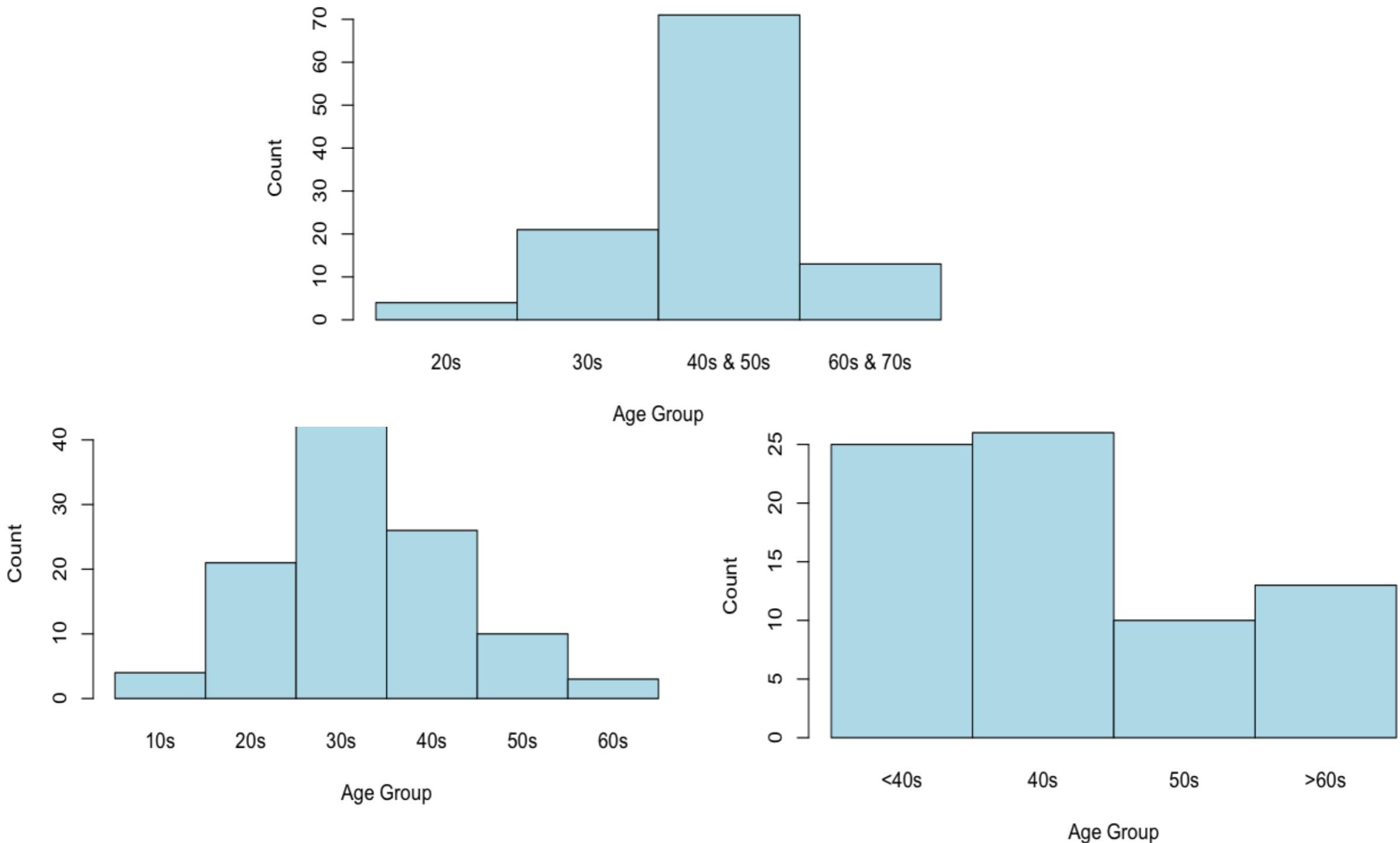
Garry Kasparov

Playing as White				Playing as Black				
8	0.10	0.09	0.15	0.19	0.14	0.11	0.05	0.05
7	0.20	0.25	0.25	0.30	0.28	0.20	0.17	0.15
6	0.24	0.26	0.47	0.48	0.42	0.57	0.29	0.28
5	0.32	0.77	0.66	1.30	1.04	0.66	0.85	0.43
4	0.77	0.67	1.51	1.99	1.57	1.06	0.68	0.72
3	0.55	0.77	1.63</td					

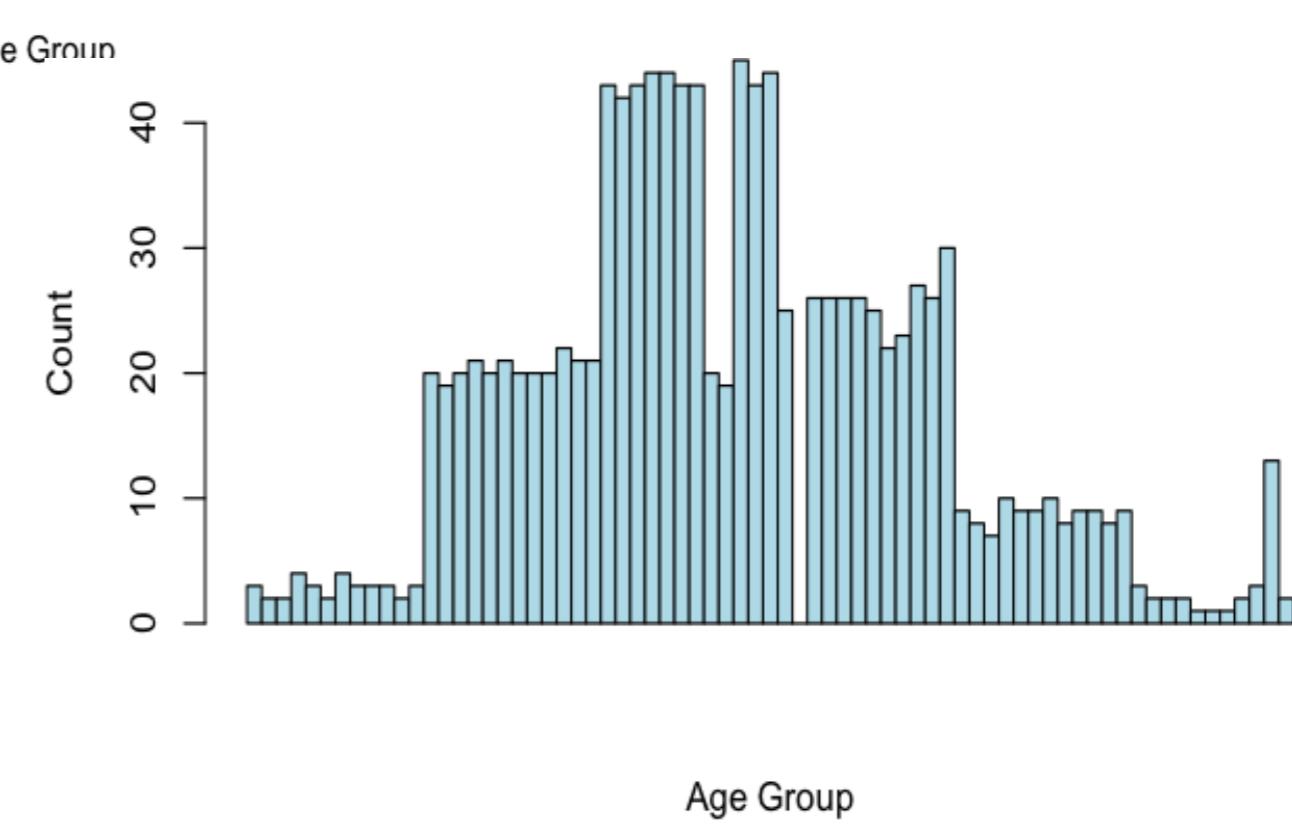
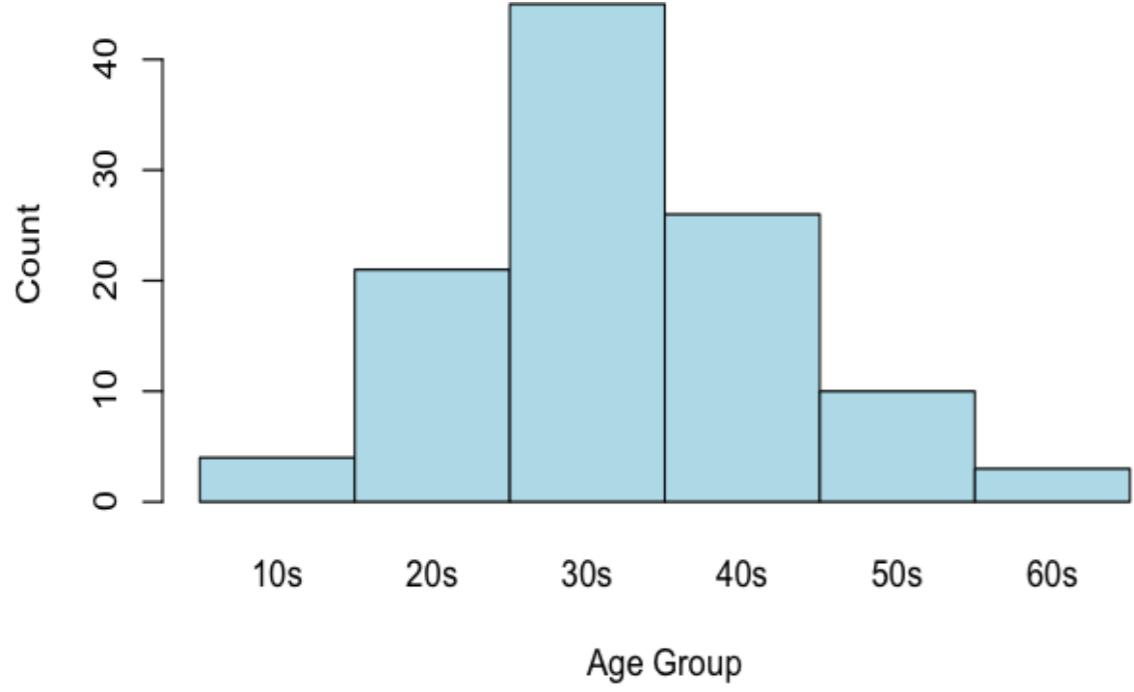
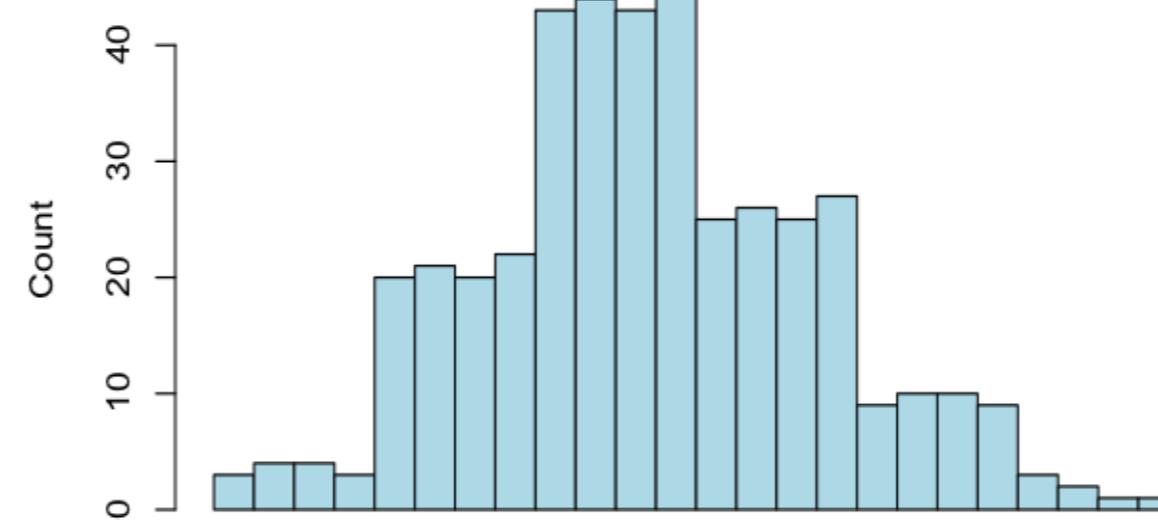
BOAS PRÁTICAS



MANTENHA OS INTERVALOS CONSISTENTES



ESCOLHA O TAMANHO APROPRIADO PARA OS INTERVALOS

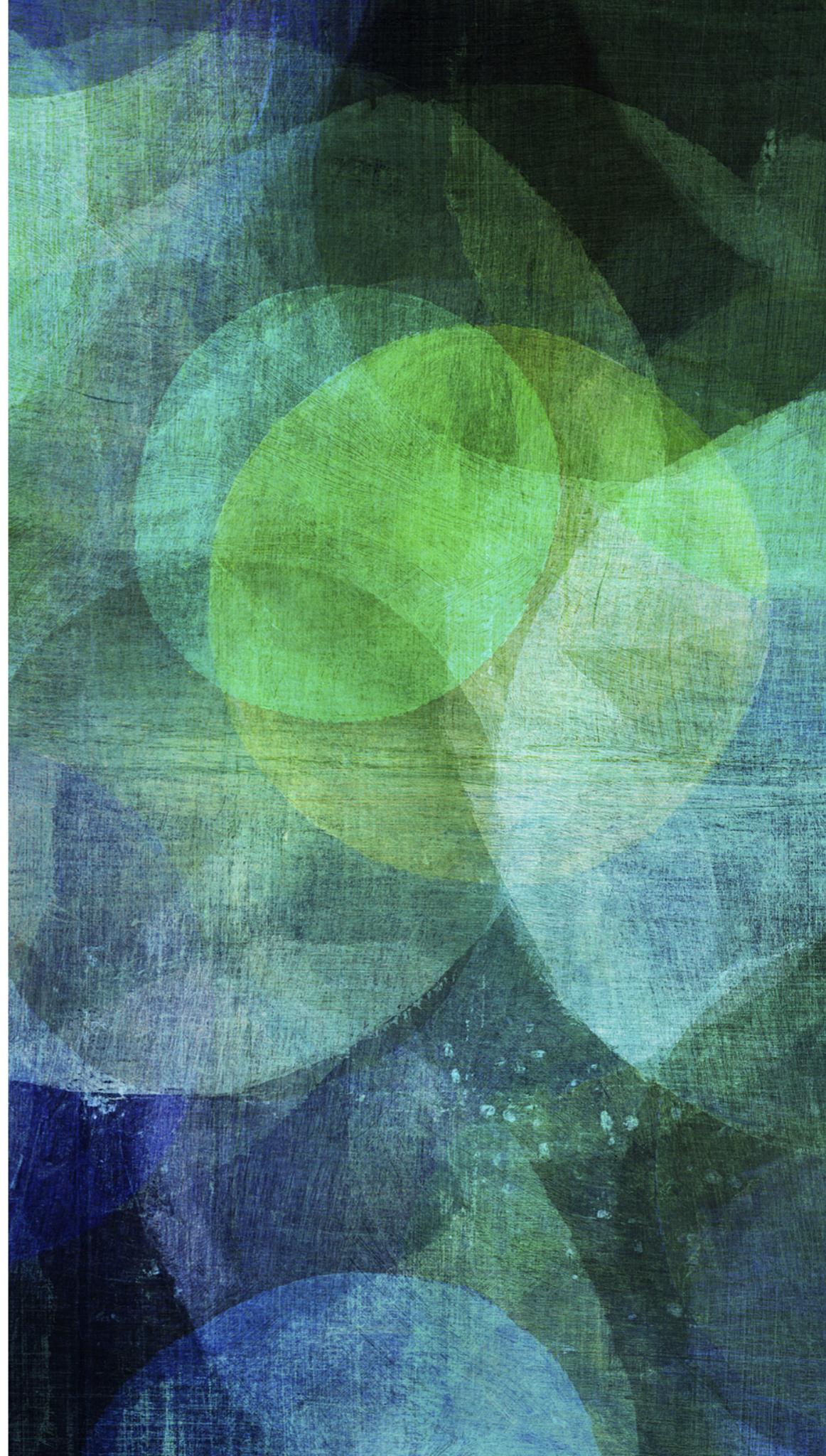


USE MÉTRICAS RESISTENTES A EXCEÇÕES

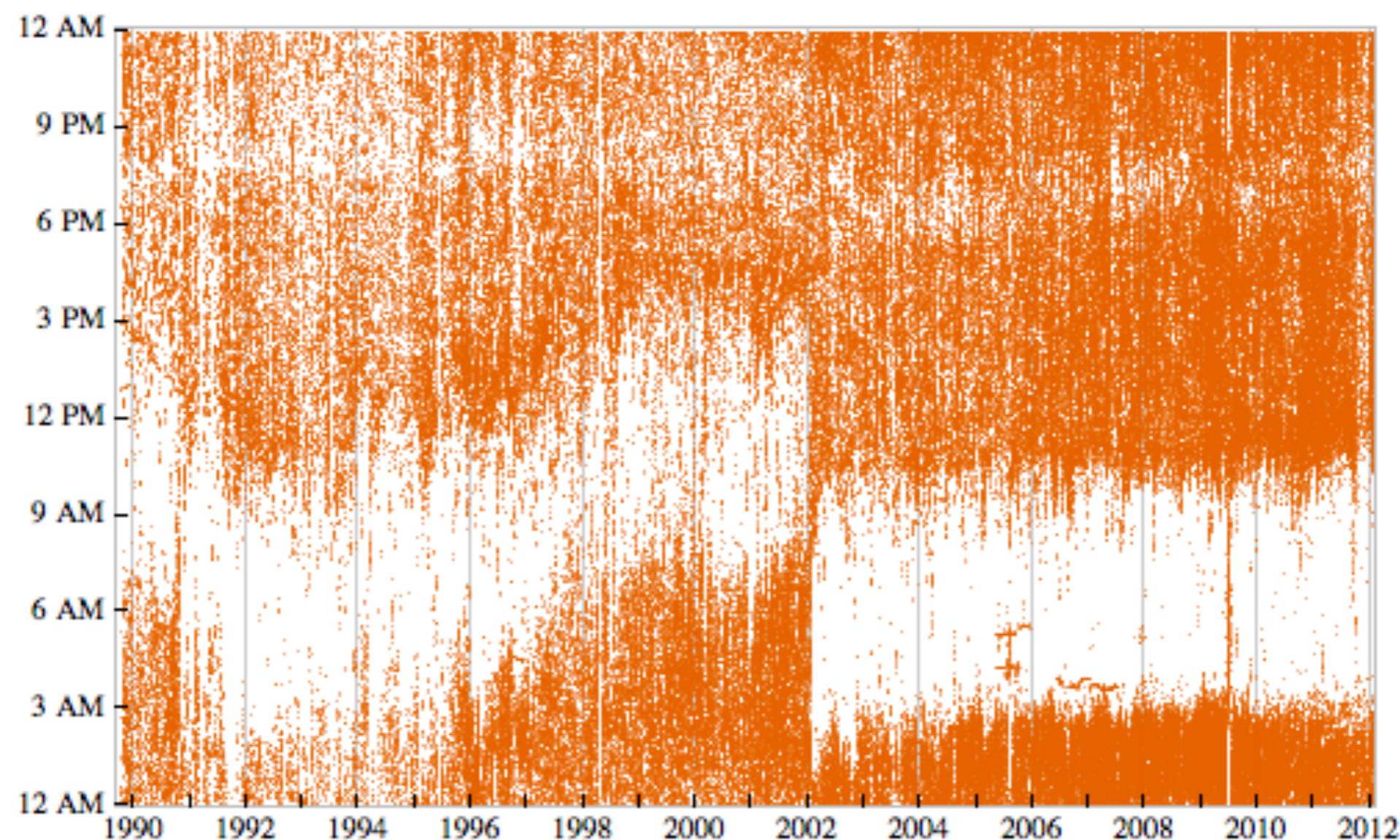
- A mediana e os percentis são medidas “resistentes” a exceções
- A média e o desvio padrão não o são

UM EXEMPLO

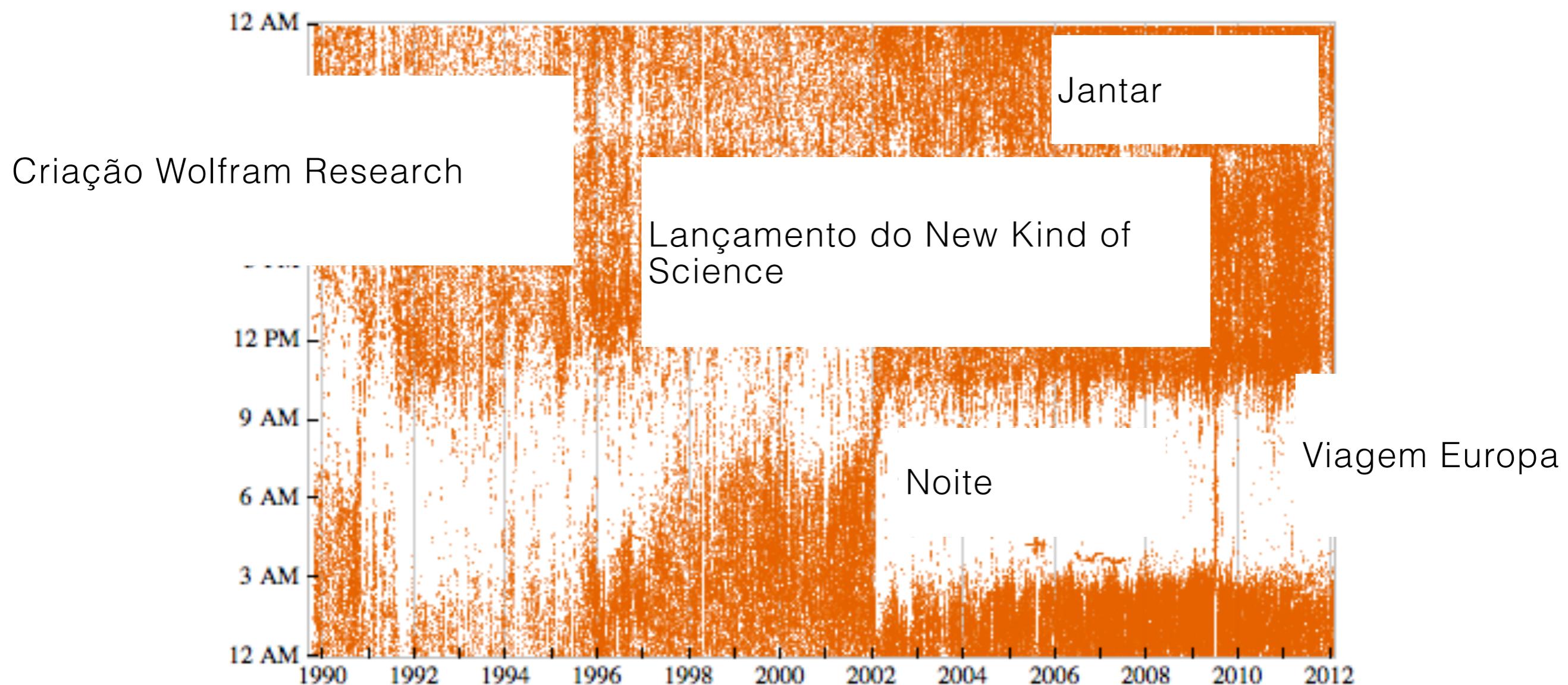
Personal Analytics



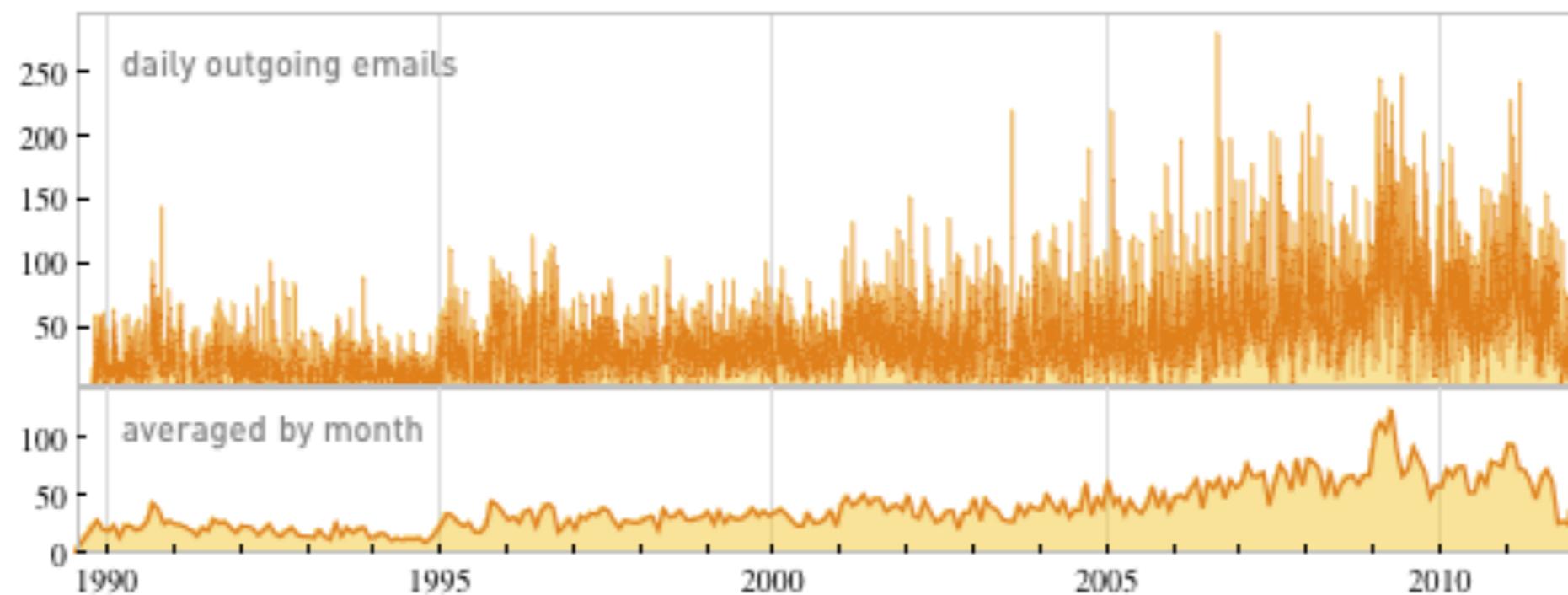
Pontos representam os cerca 300.000 **e-mails enviados** desde 1989



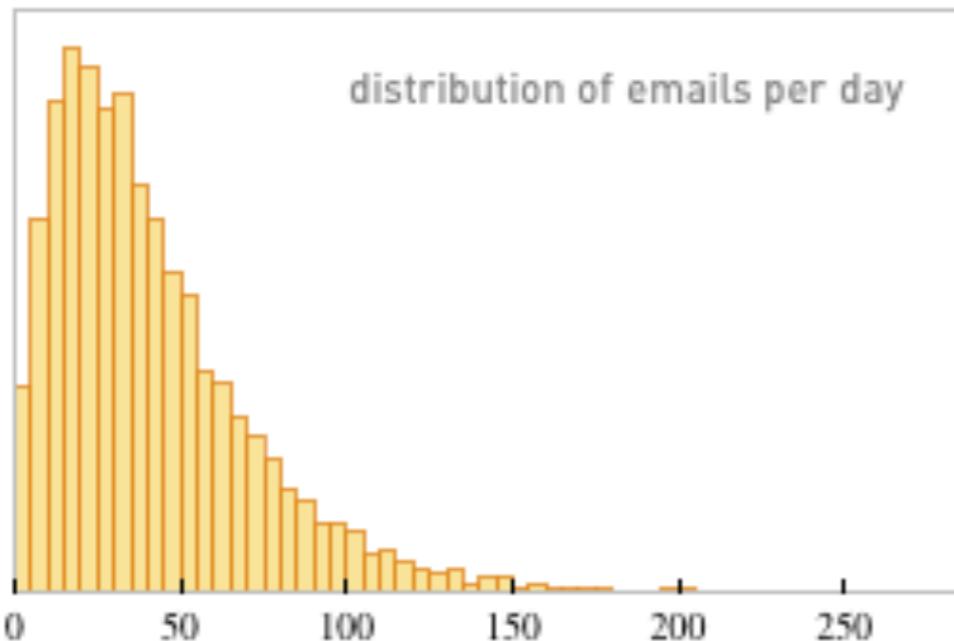
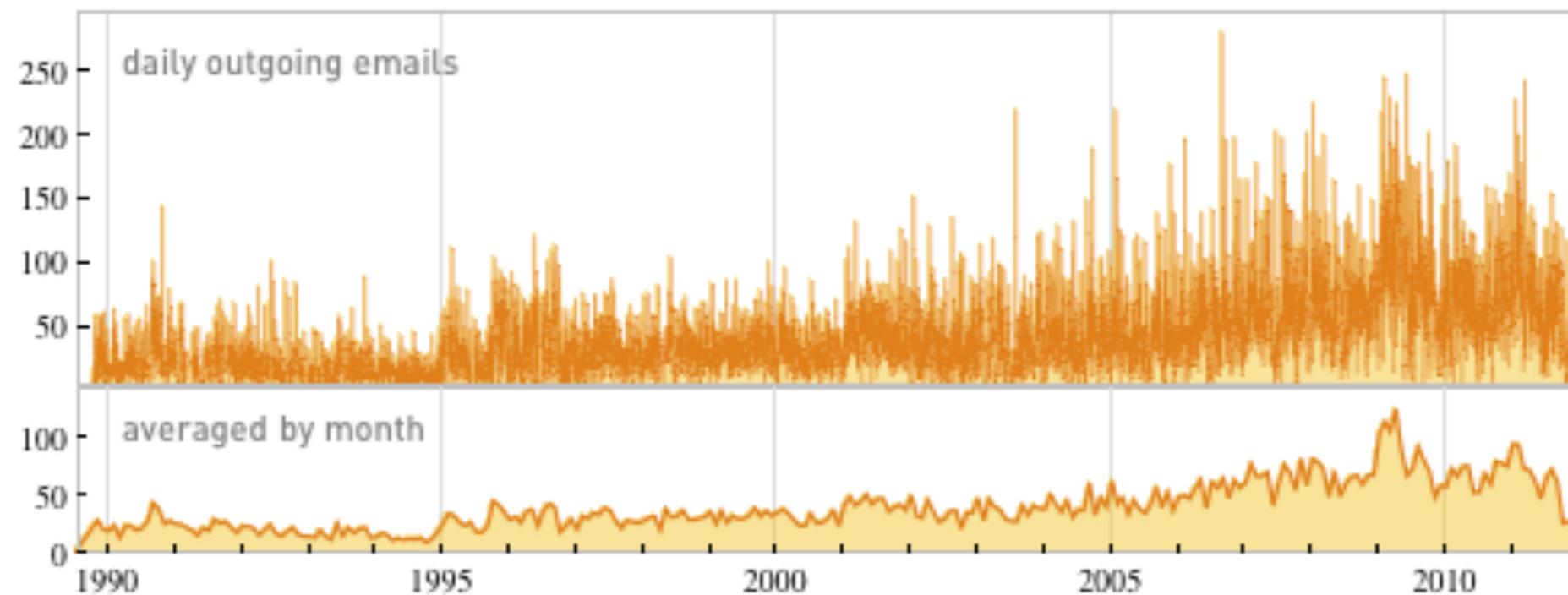
Pontos representam os cerca 300.000 **e-mails enviados** desde 1989



Quantidade de **e-mails enviados** ao longo do tempo e sua média mensal



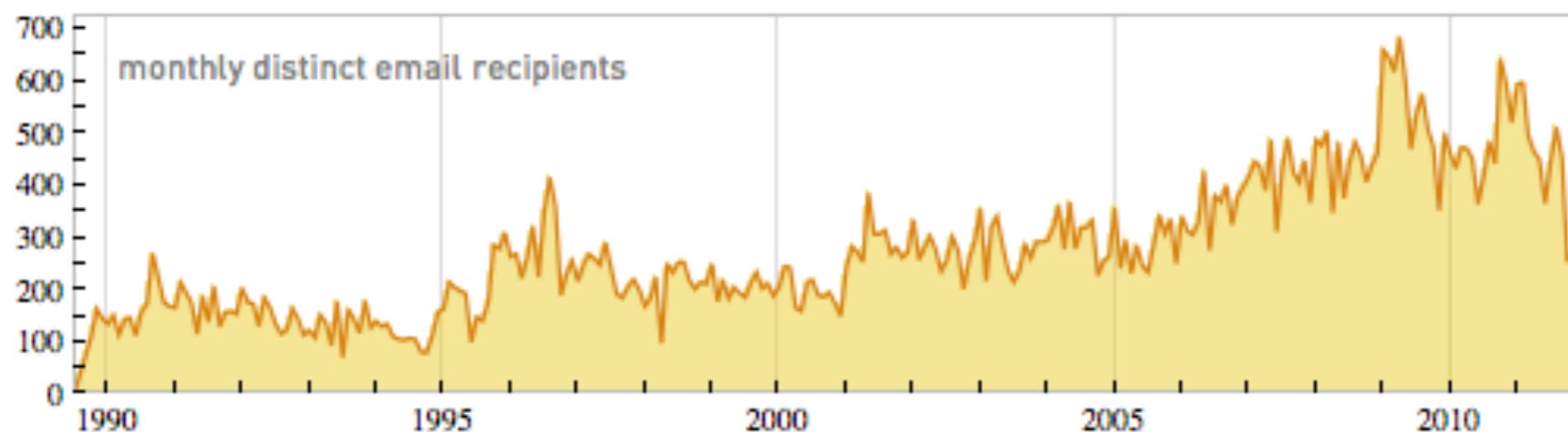
Quantidade de **e-mails enviados** ao longo do tempo e sua média mensal



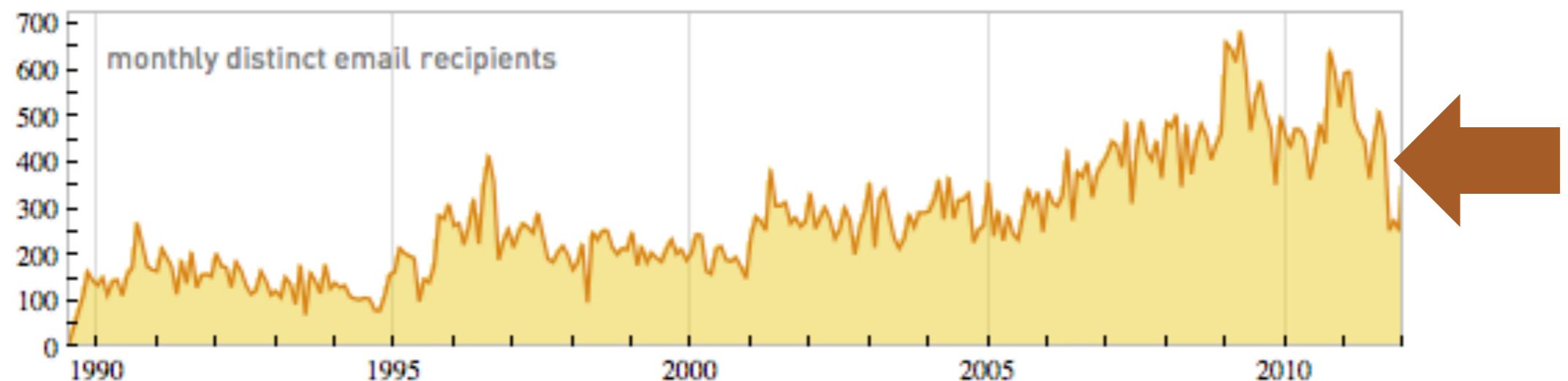
Mudança de perspectiva:
distribuição?

A cauda parece seguir uma lei de
potência

Número de **diferentes destinatários** por dia

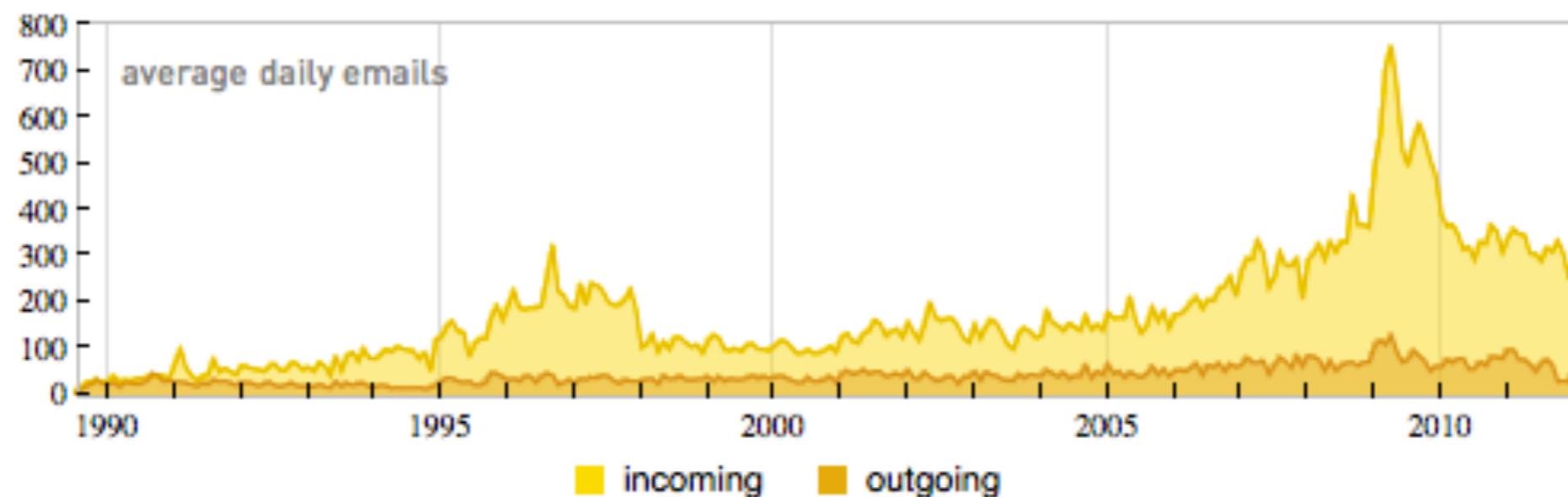


Número de **diferentes destinatários** por dia



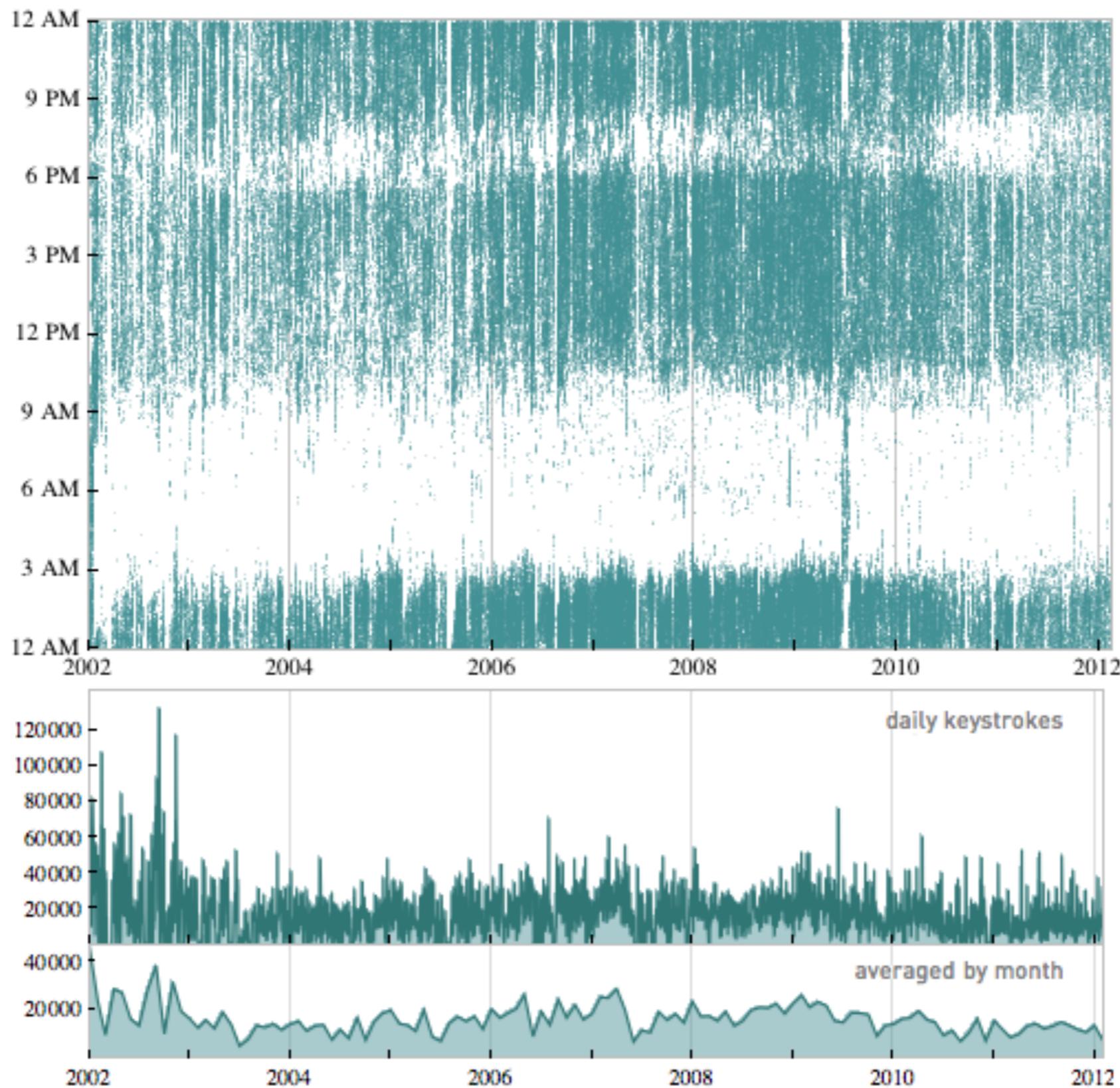
Picos coincidem com inícios de projetos, quando
esta normalmente envolvido com um número
maior de pessoas

Número de **diferentes remetentes** por dia

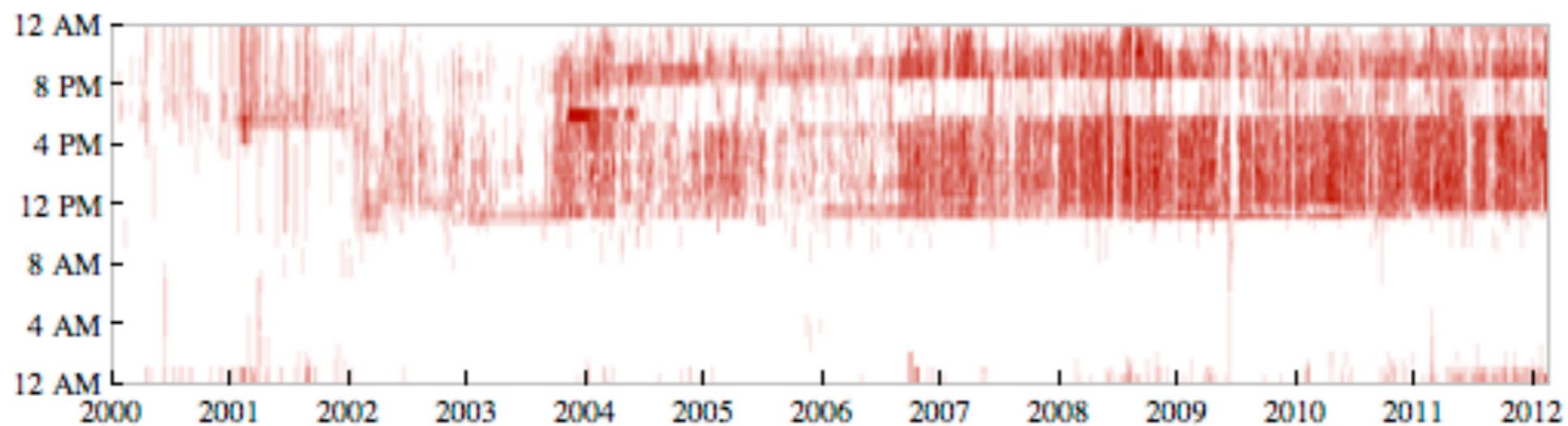


Picos coincidem com lançamentos de
projetos: Mathematica 3 e Wolfram|
Alpha

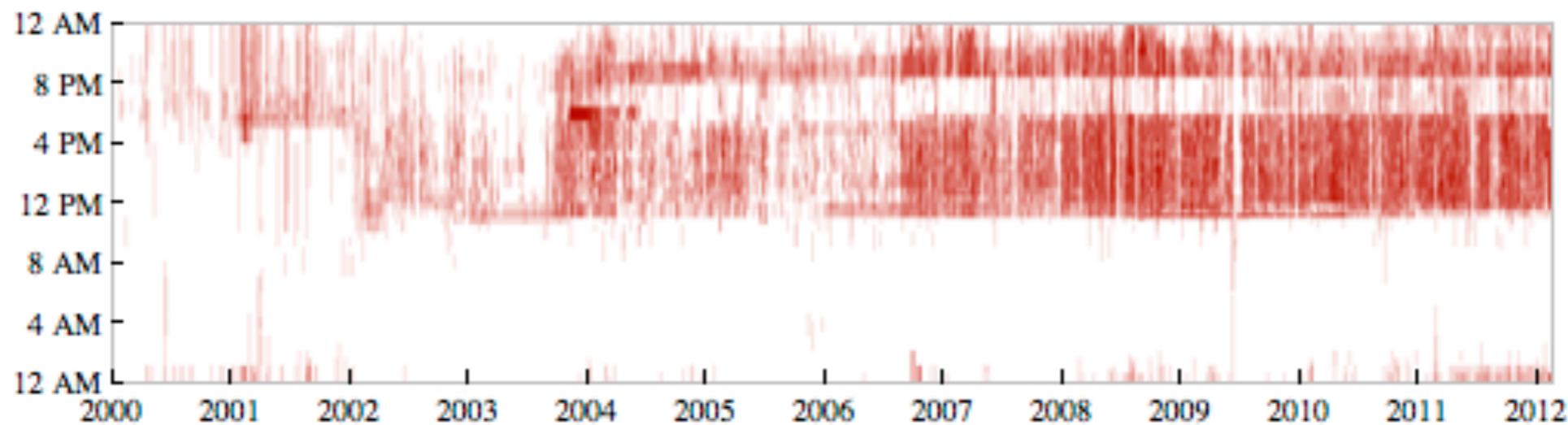
Cerca de 100.000.000 de **teclas** que digitou



Eventos na agenda



Eventos na agenda

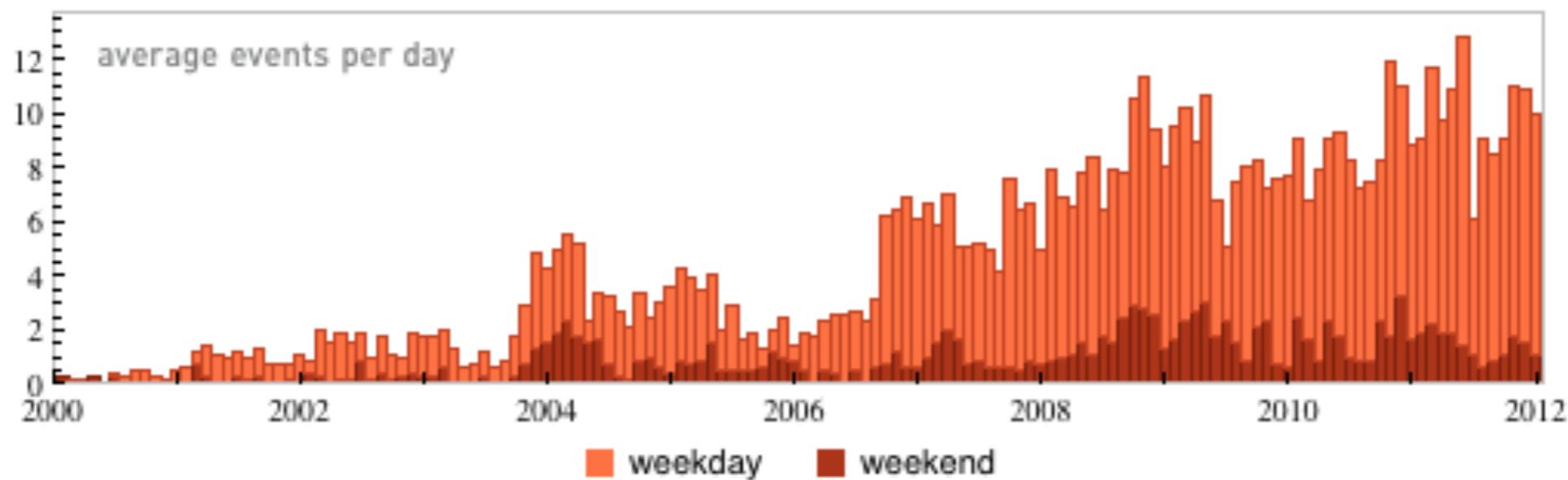


Antes de 2002, trabalhava sozinho na escrita do NKS

Depois o engajamento na empresa fez com que tivesse mais compromissos

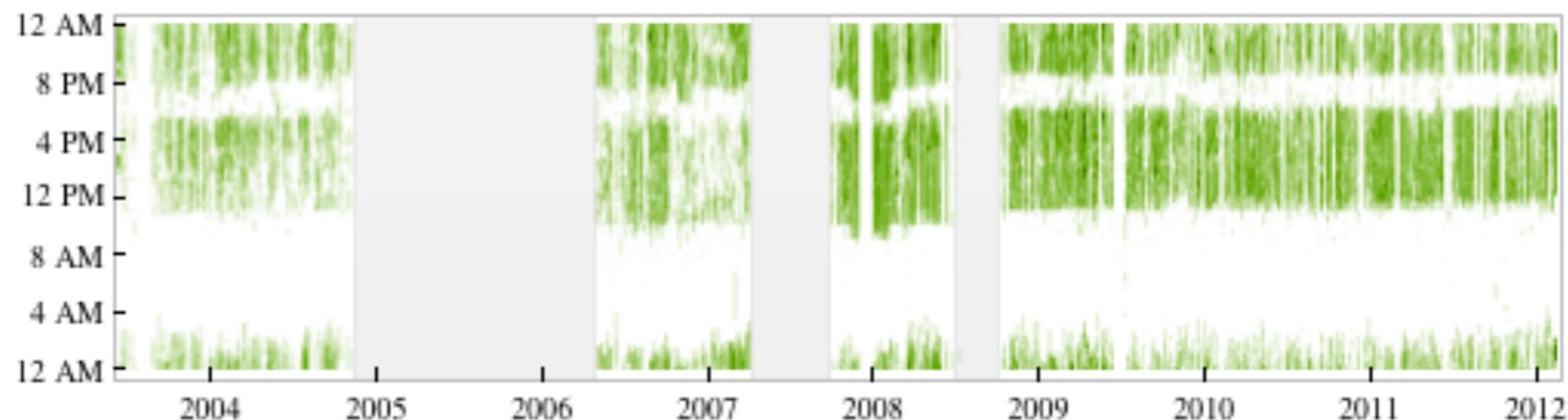
Observe a faixa do período de jantares

Distribuição do número de **eventos** por dia



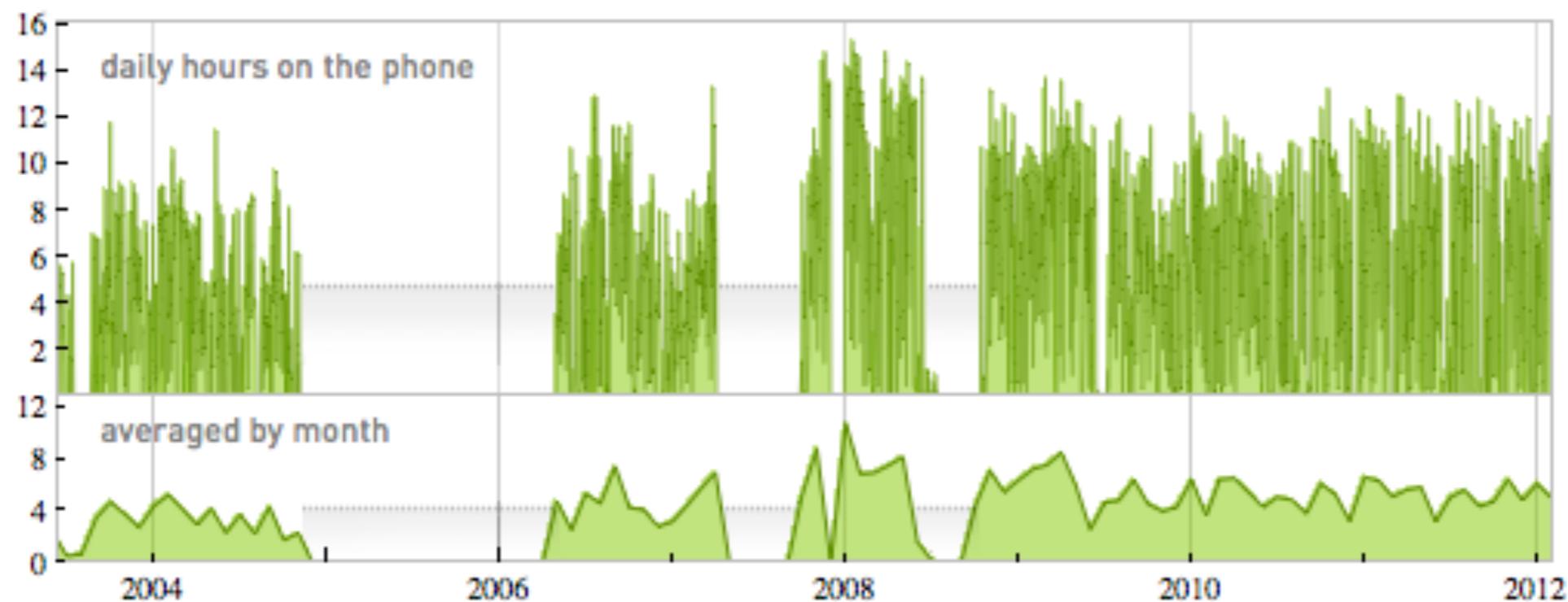
Mudança de paradigma: começou a trabalhar melhor “em público”. Antes ele era um CEO “remoto”

Chamadas telefônicas ao longo dos dias

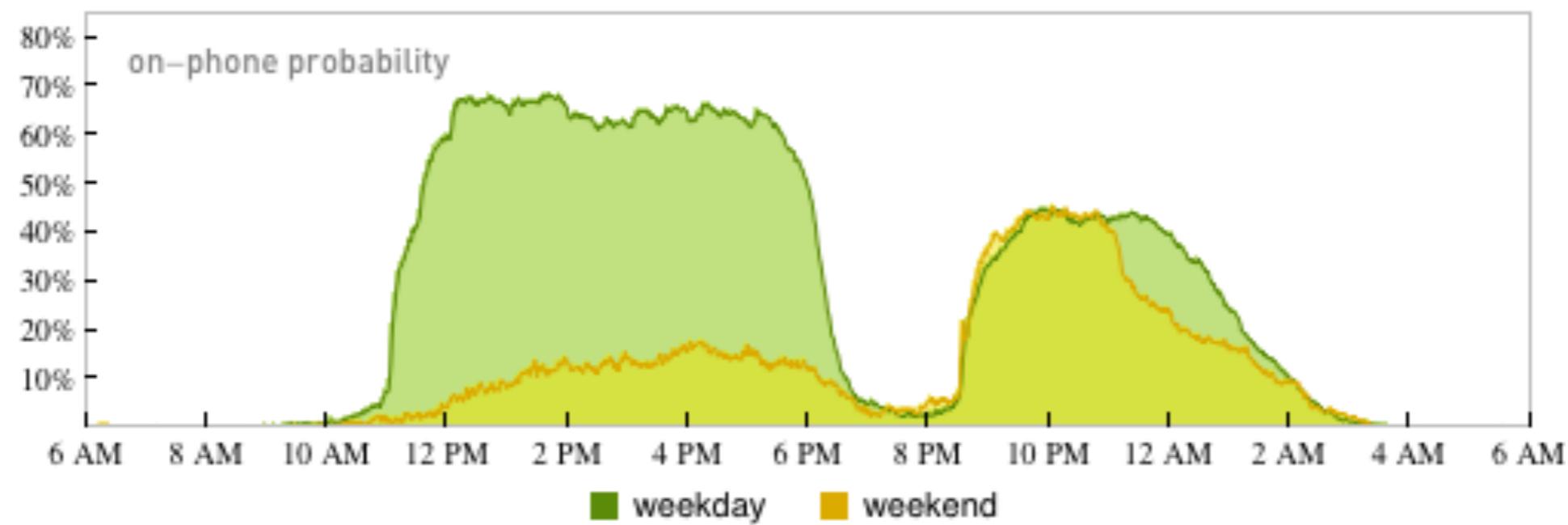


Faixas em cinza indicam ausência de dados

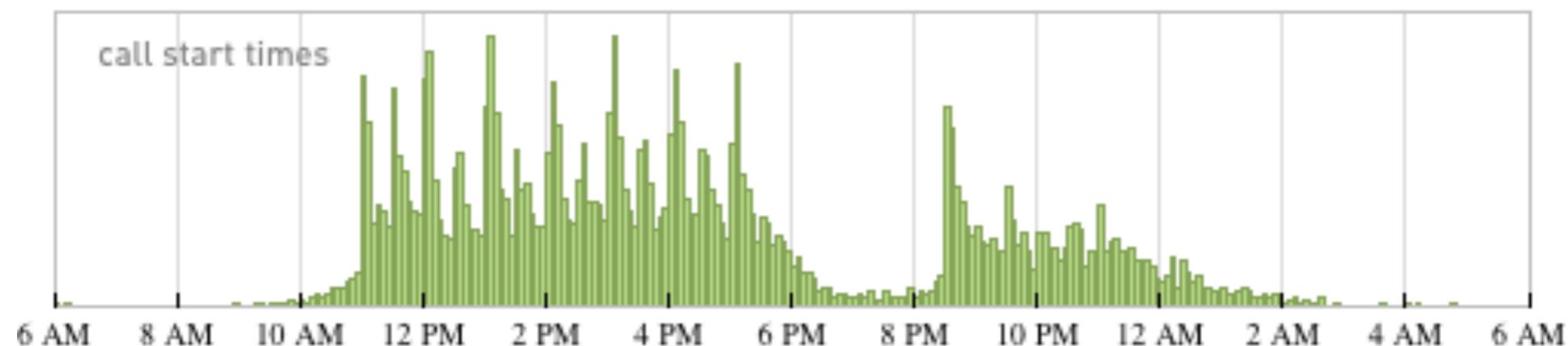
Sim, ele passa muitas horas do dia ao telefone!!!



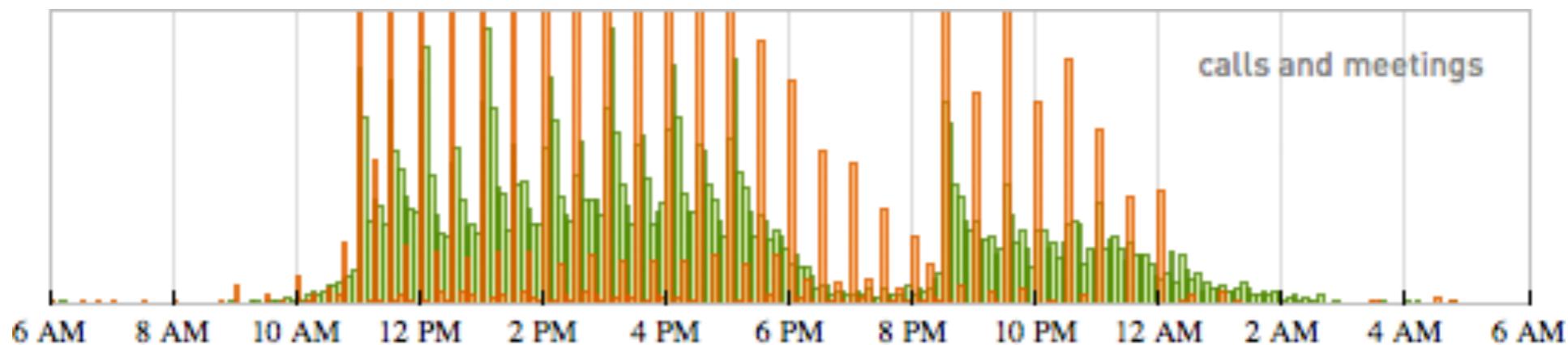
Veja a probabilidade de ele estar ao telefone nas diversas horas do dia!



Número de chamadas que se iniciam ao longo do dia



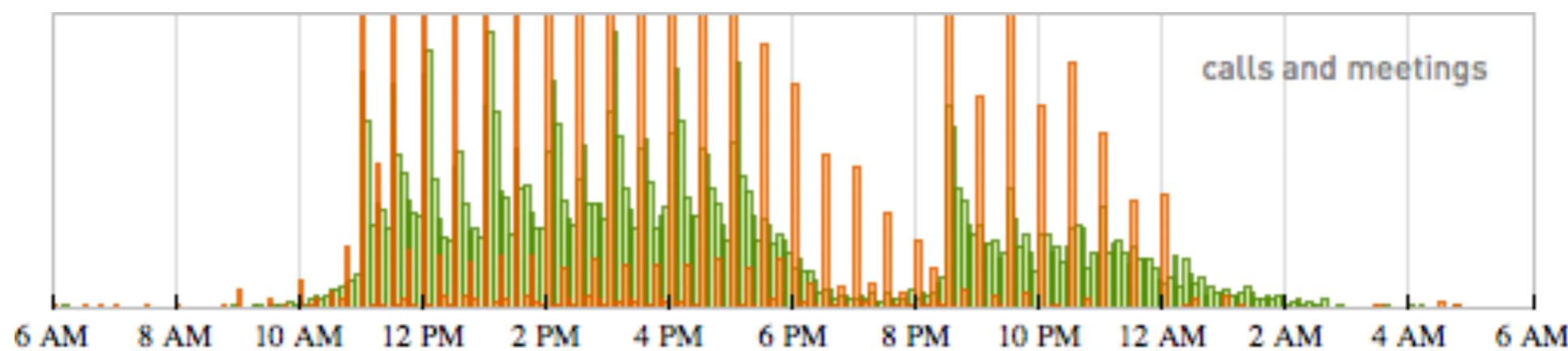
Número de chamadas que se iniciam ao longo do dia



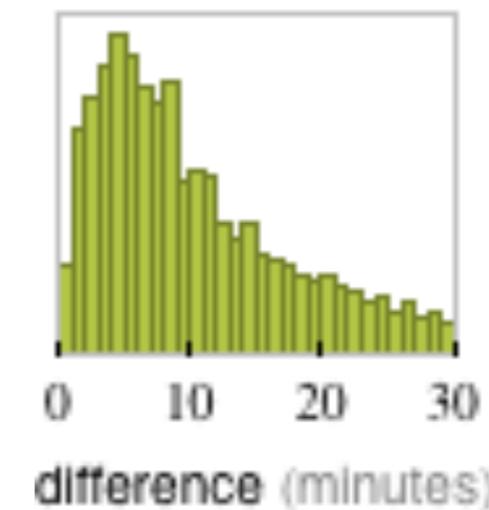
Número de chamadas e eventos na agenda

Há alguma correlação?

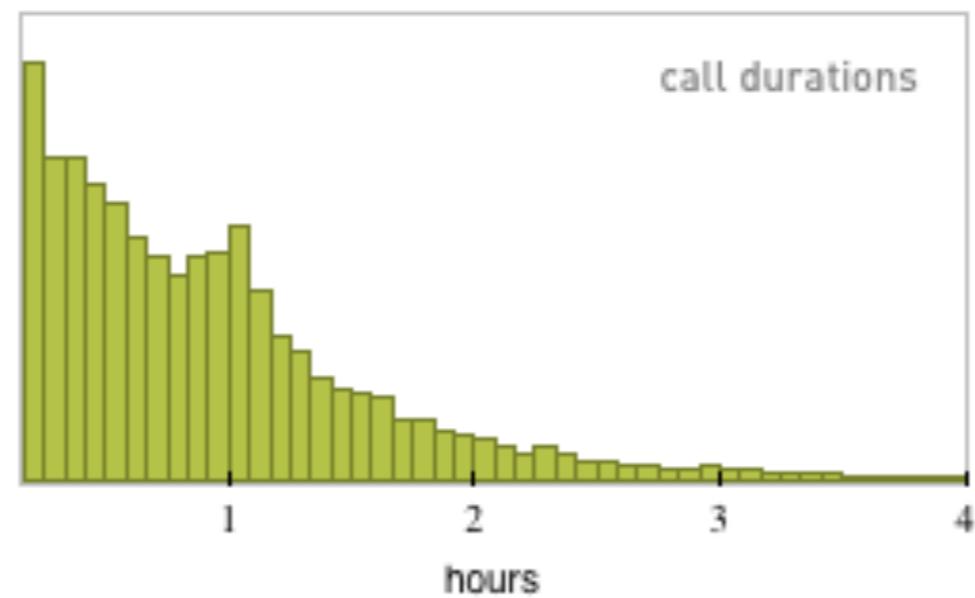
Por que?



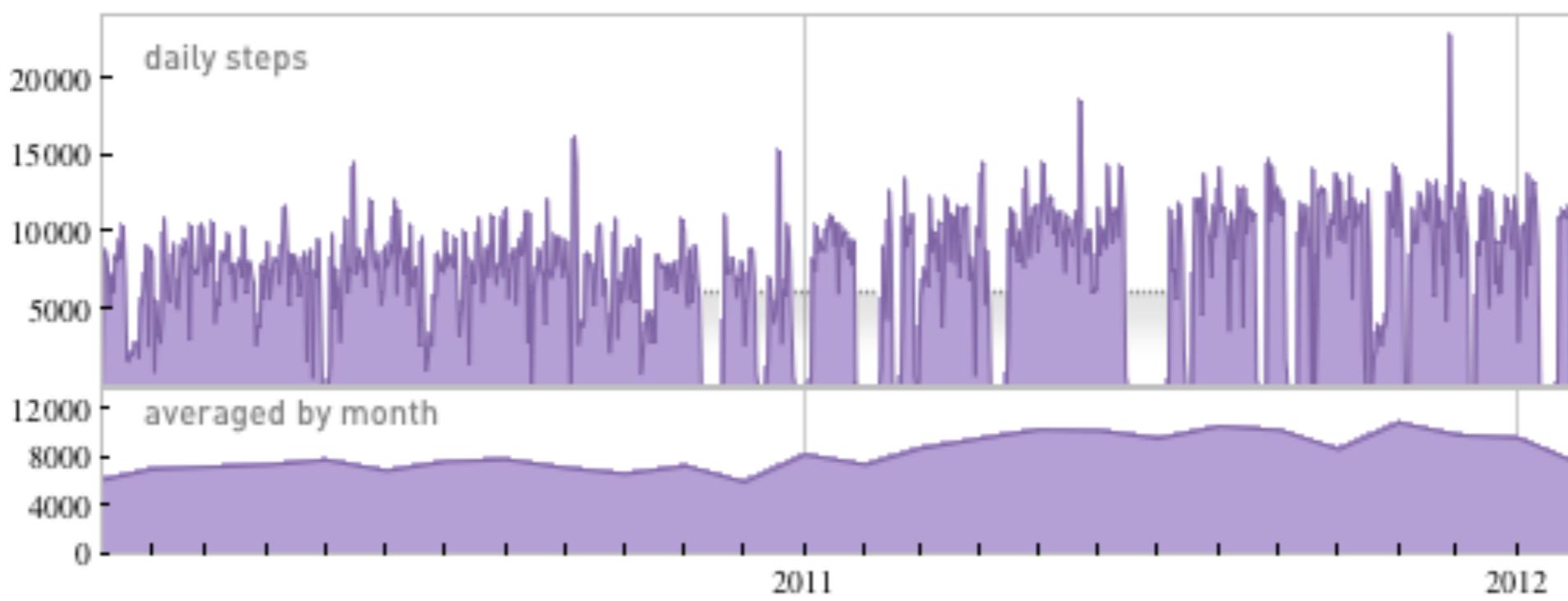
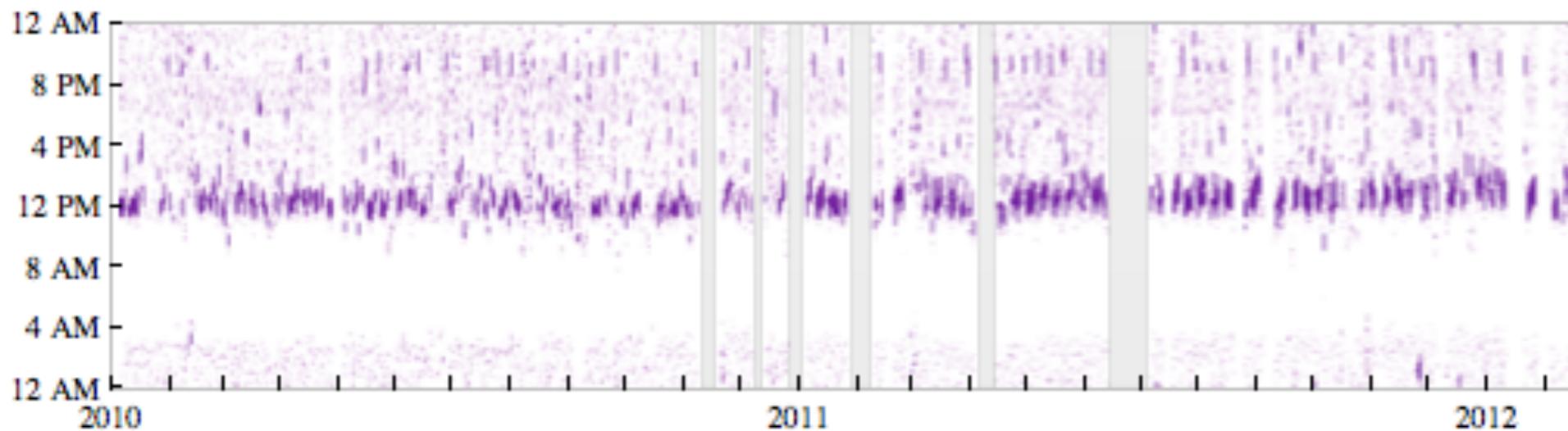
Reuniões começam com 2 minutos de **atraso**



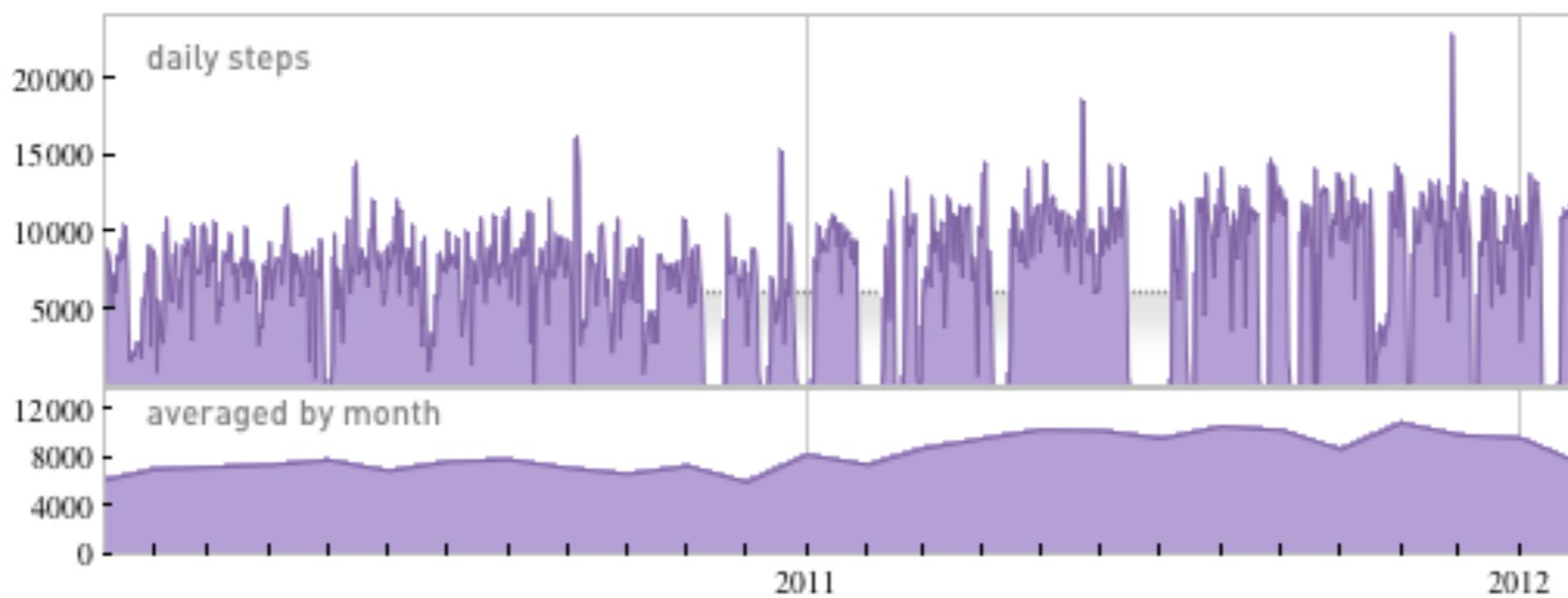
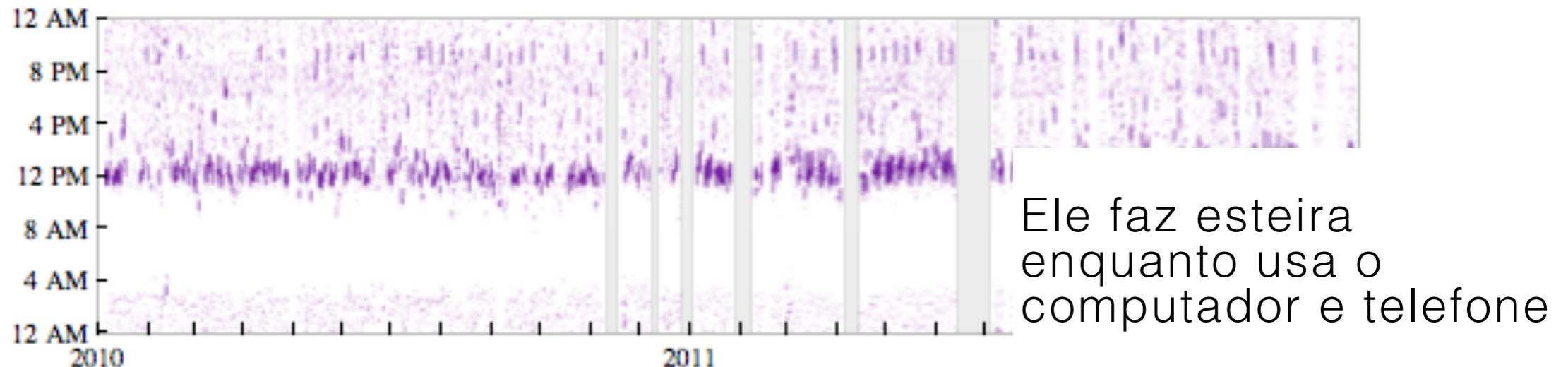
Distribuição das durações das chamadas
Por que há um pico em 1 hora?

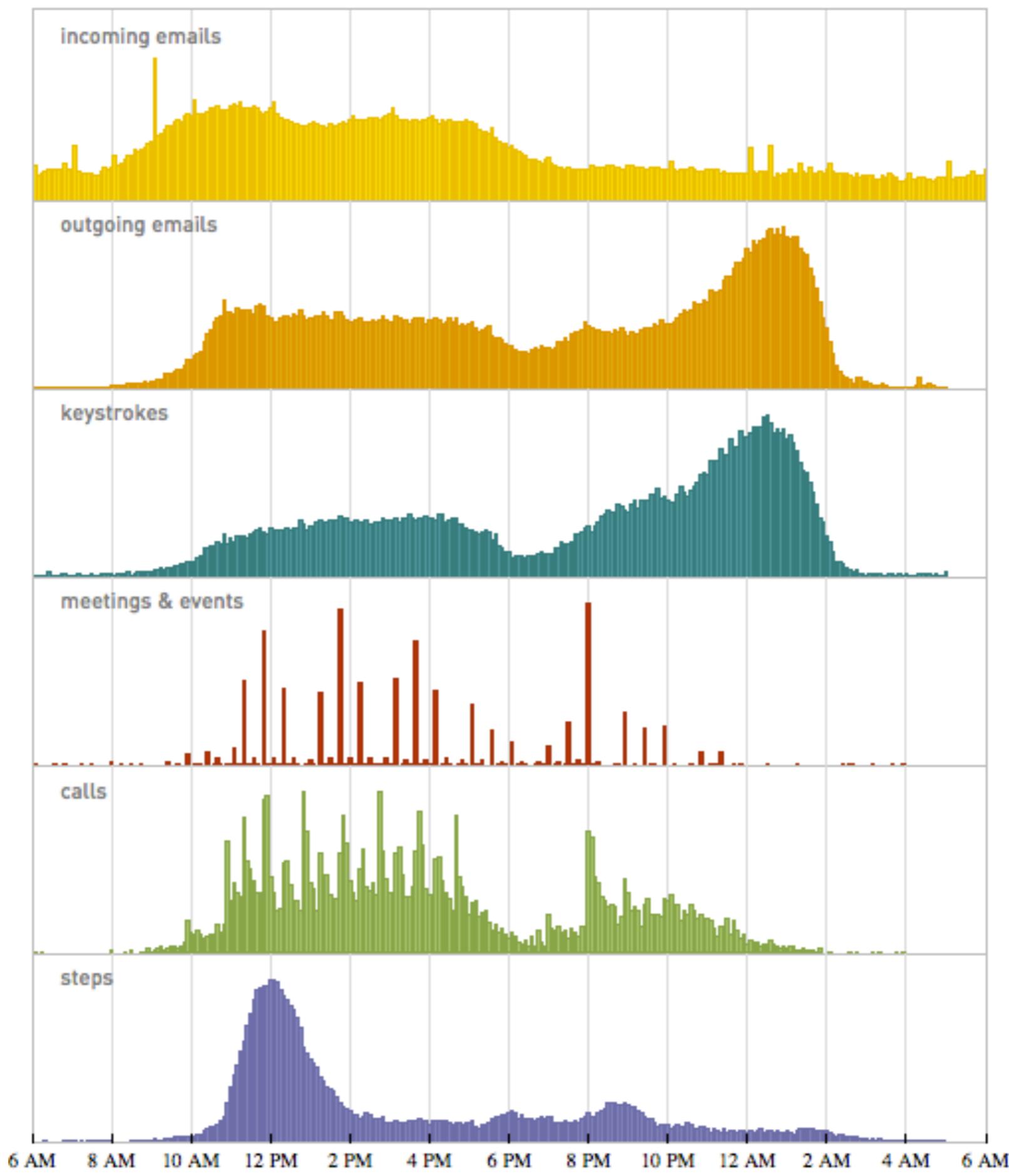


Ele contou até quantos **passos** ele deu!! Credo!



Ele contou até quantos **passos** ele deu!! Credo!

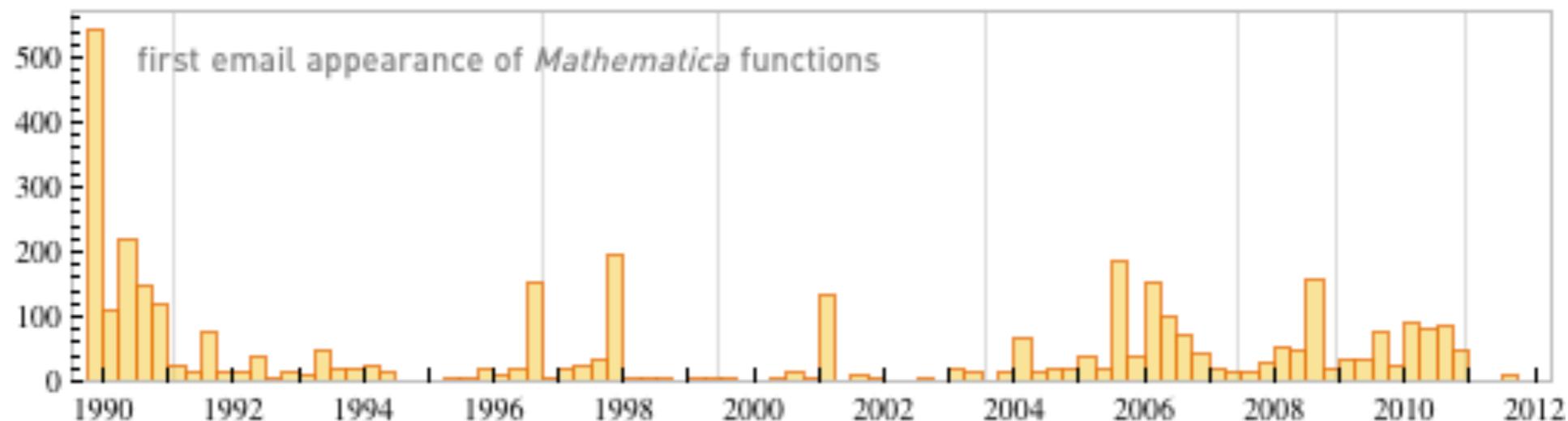




É possível correlacionar os **hábitos** com os **momentos criativos**?

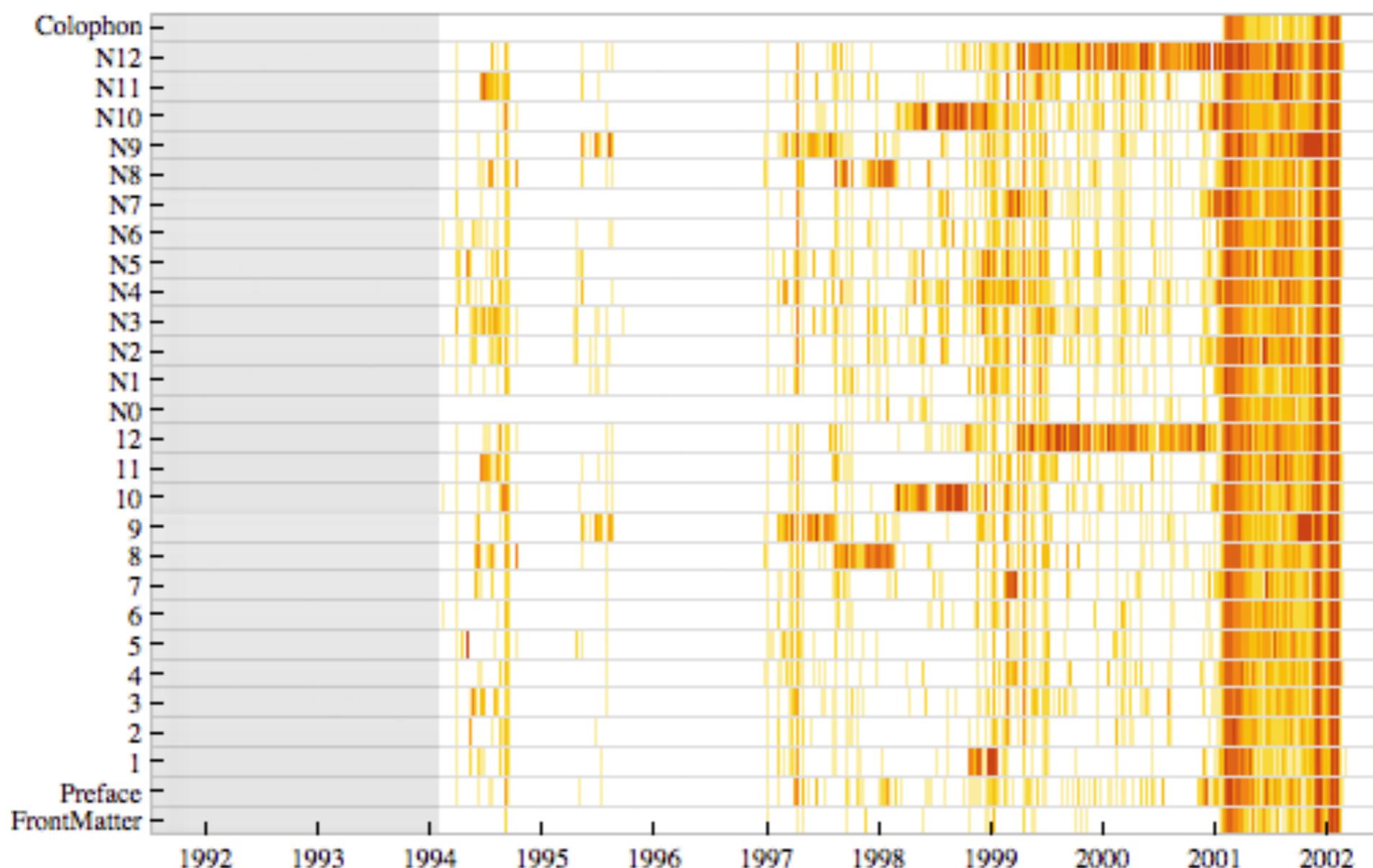
Analisou as “**novas**” **palavras** que apareciam em seus e-mails

Cerca de 33.000 diferentes palavras na última década



Tentativa de visualização do progresso criativo

Modificações no texto de NKS



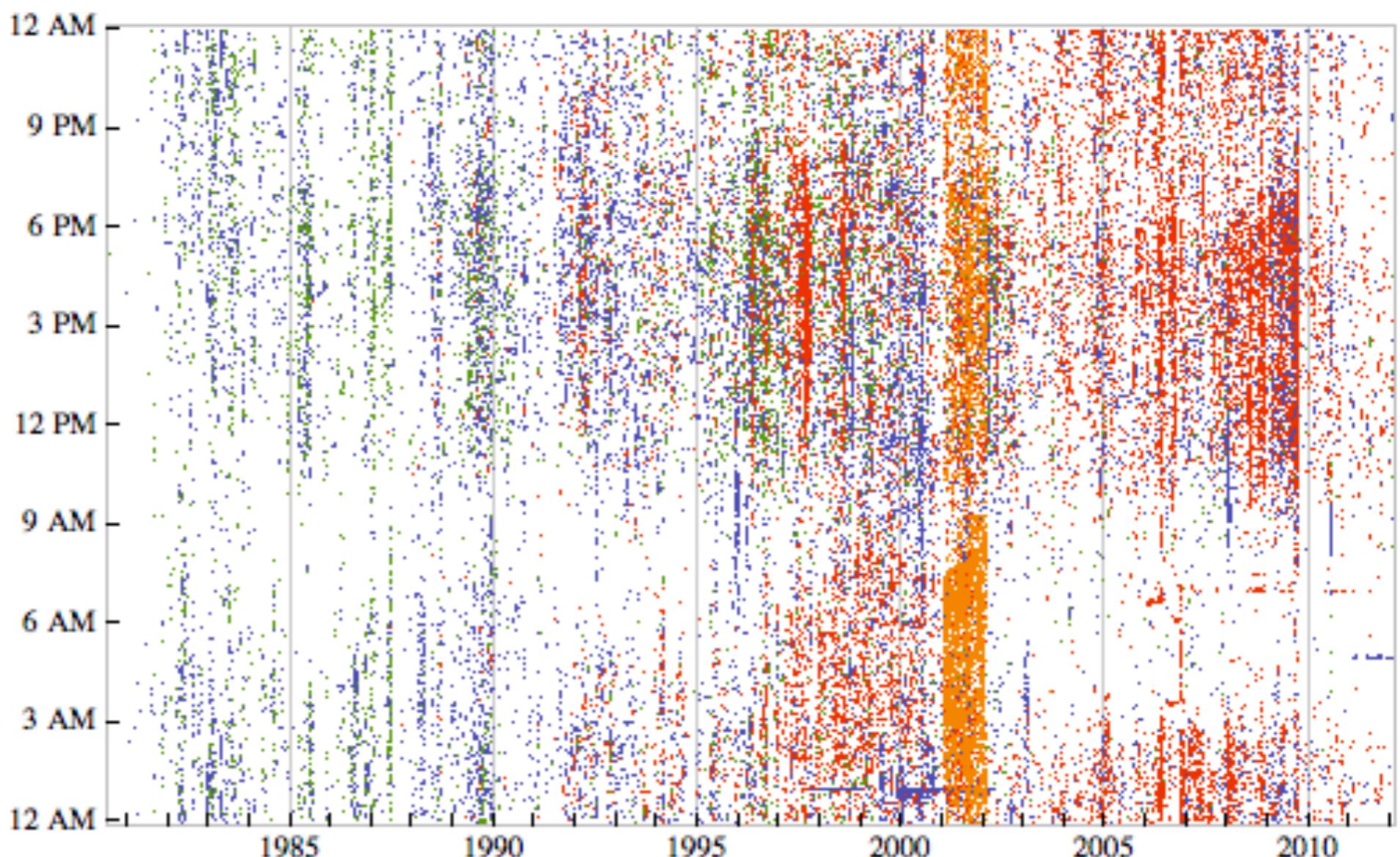
Modificações de todos os seus **arquivos**

Azul: texto

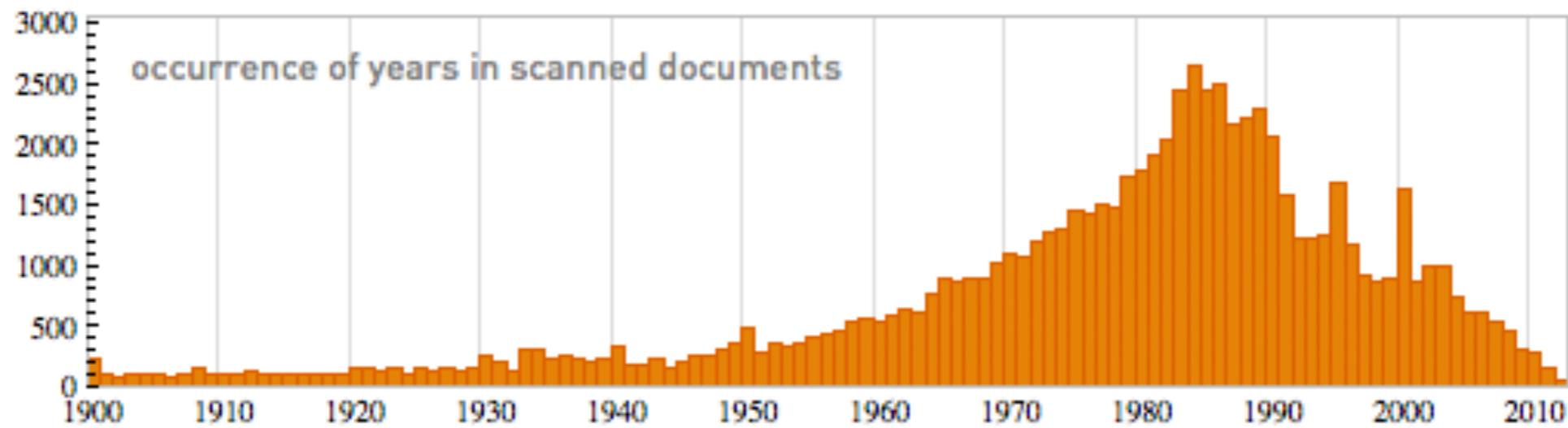
Verde: códigos em C

Vermelho: Mathematica

Laranja: leiautes



Ocorrência de **anos em documentos em papel, digitalizados**



A propósito, ele ainda tem muitos dados não analisados: seu genoma completo, rastreamento via GPS, monitoramento de mudanças de sala, registros médicos...

<http://blog.stephenwolfram.com/2012/03/the-personal-analytics-of-my-life/>