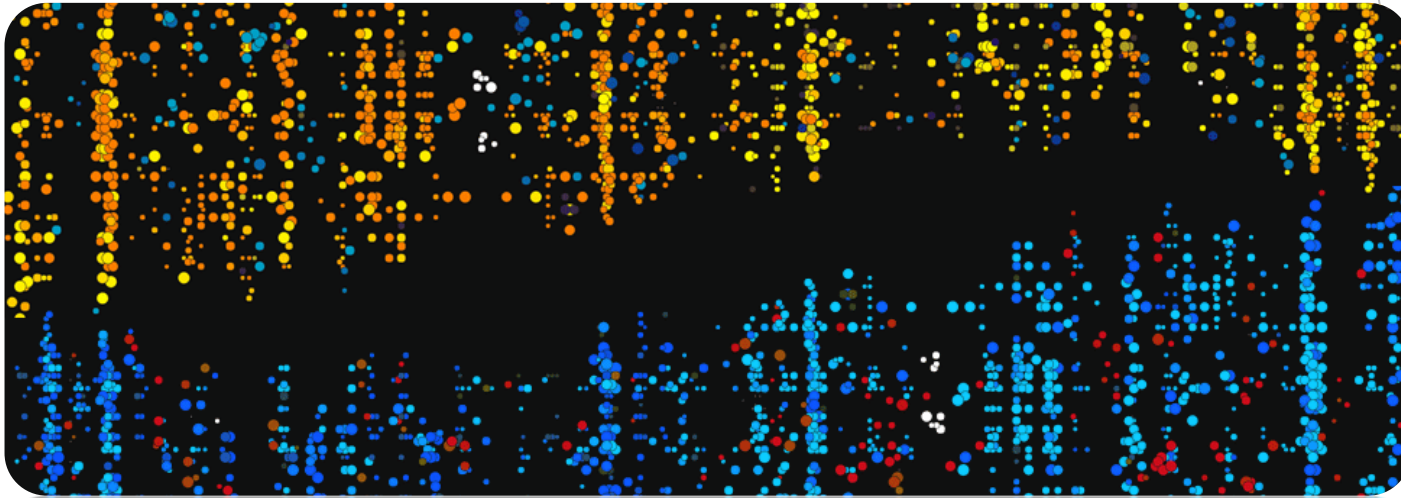


Data Science com Streamlit



source: <http://mkweb.bcgsc.ca/images/masthead/erica-savig-cytometry.png>

André Filipe M. Batista, Ph.D.

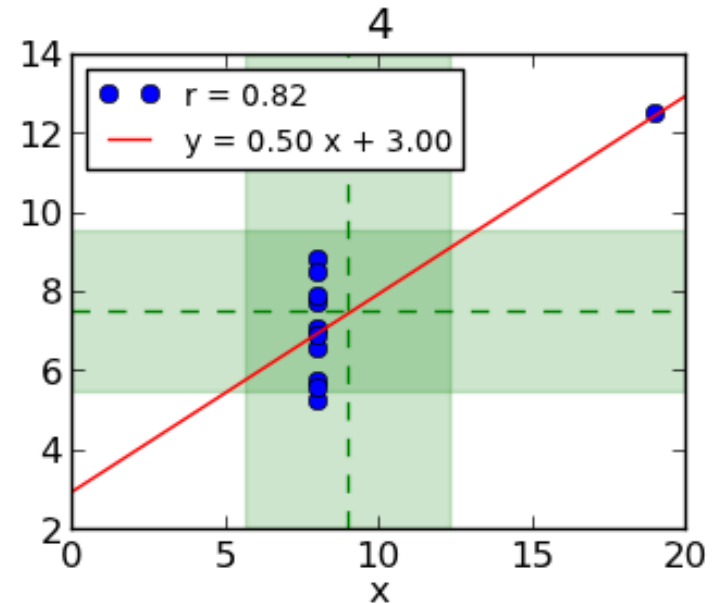
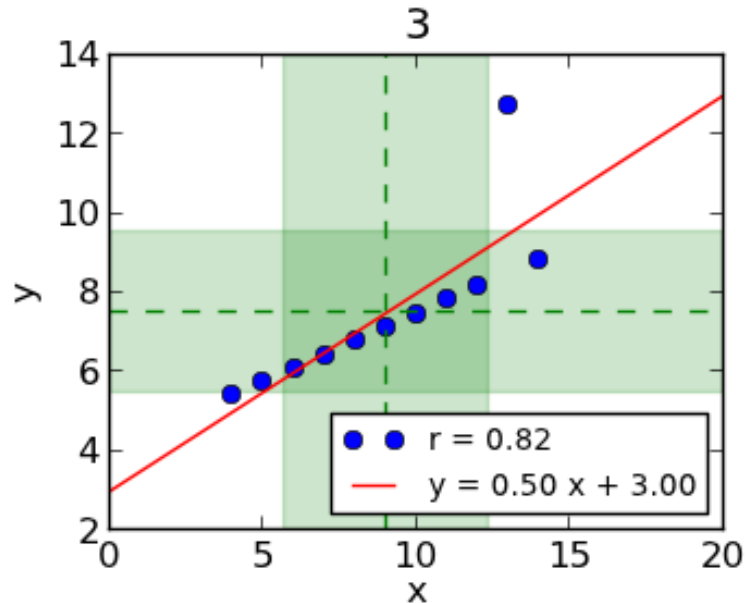
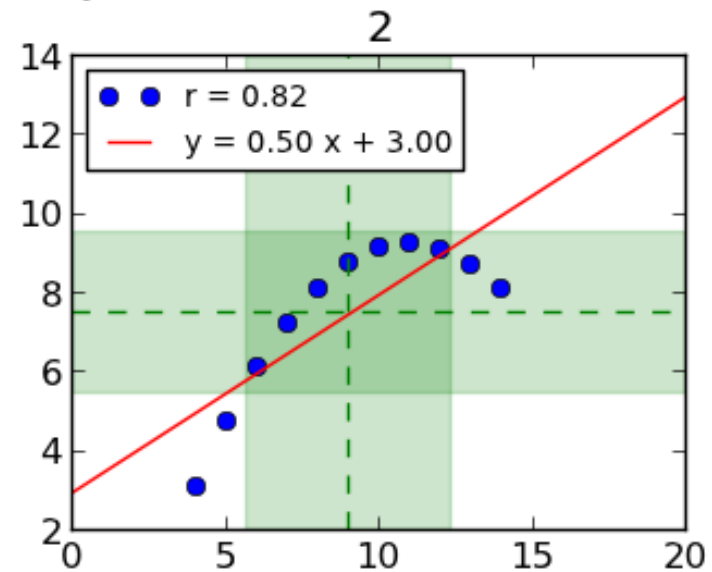
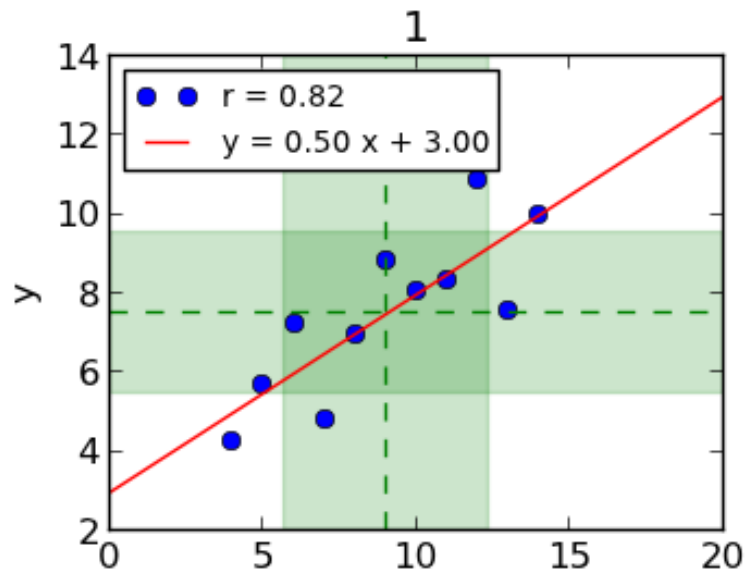
2020

Observe o conjunto de dados abaixo:

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

Agora visualize-os:

Anscombe's Quartet



O que é visualização?

Representação
visual de dados
com foco em
reforçar a
cognição
humana

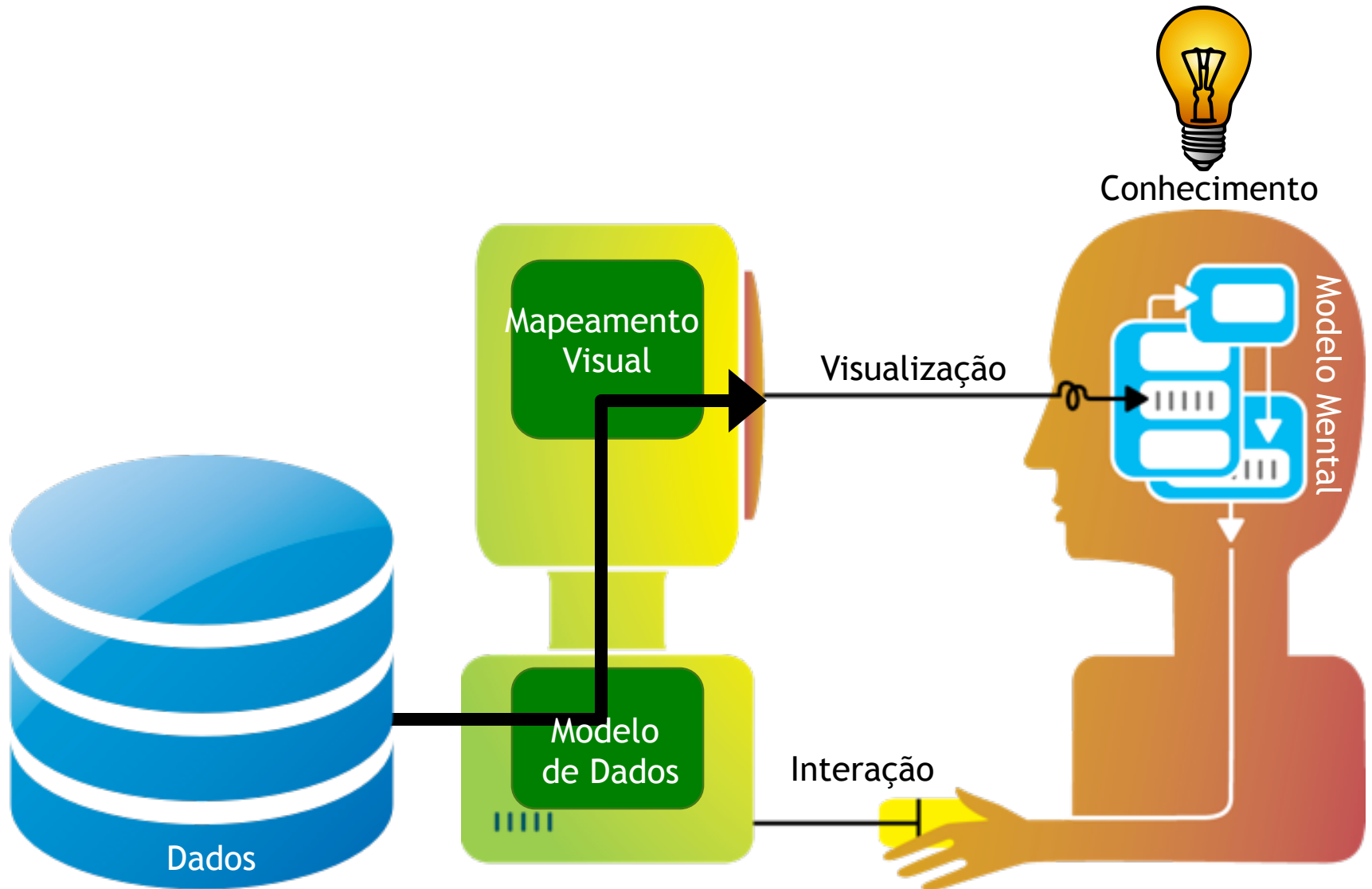


http://www.infovis-wiki.net/index.php/Information_Visualization

A importância da Visualização

- Permite diminuir a carga de memória para relembrar sobre conceitos/fatos
- Reduz o tempo de busca sobre conceitos
- Nos permite realizar inferências perceptuais

Dados e Visualização



Visual Analytics (VA)



**“The science of analytical reasoning
facilitated by interactive visual interfaces”¹**

¹Thomas, James J., and Kristin A. Cook, eds. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005.

O que é Visual Analytics?

Visualização +

- Representação de dados
- Interação e análise
- Disseminação (reprodutibilidade)



Por que Visual Analytics?

- Nós estamos coletando e gerando dados em uma velocidade superior a que podemos absorver tal volume
- Estes dados normalmente são complexos, ambíguos e com ruídos
 - Não é apenas pelo fato de ser Big Data, mas pelo fato de que lidar com grandes volumes de dados requer a interpretação e a compreensão humana
 - Ou melhor:
- Grande problema: “*humans don’t scale*”

Interatividade

1. Selecionar: realçar algo de interesse
2. Explorar: exibir algo a mais
3. Reconfigurar: exibir os dados de uma maneira diferente
4. Codificar: exibir diferentes representações visuais
5. Abstrair/Elaborar: exibir mais ou menos detalhes
6. Filtrar: Exibir algo válido para uma condição
7. Conectar: Exibir itens relacionados

Yi, J. S., ah Kang, Y., Stasko, J. T., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1224-1231.

Baby Name Wizard

Baby Name >

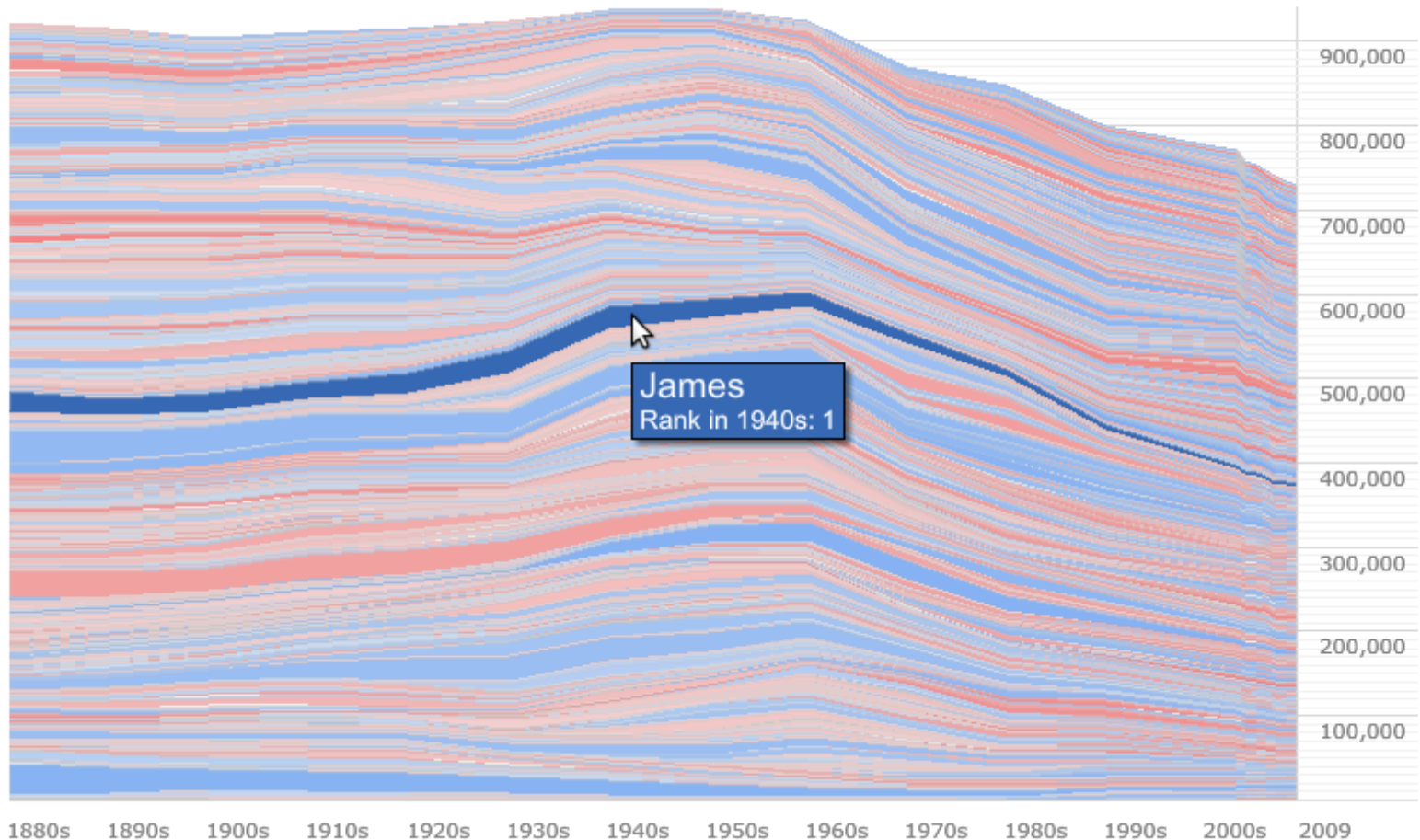
● Both ○ Boys ○ Girls

2009 rank: boys

1000	500	100	25	1
------	-----	-----	----	---

girls

1000	500	100	25	1
------	-----	-----	----	---



Click a name graph to view that name. Double click to read more about it.

Baby Name Wizard – Filter by “James”

Baby Name >

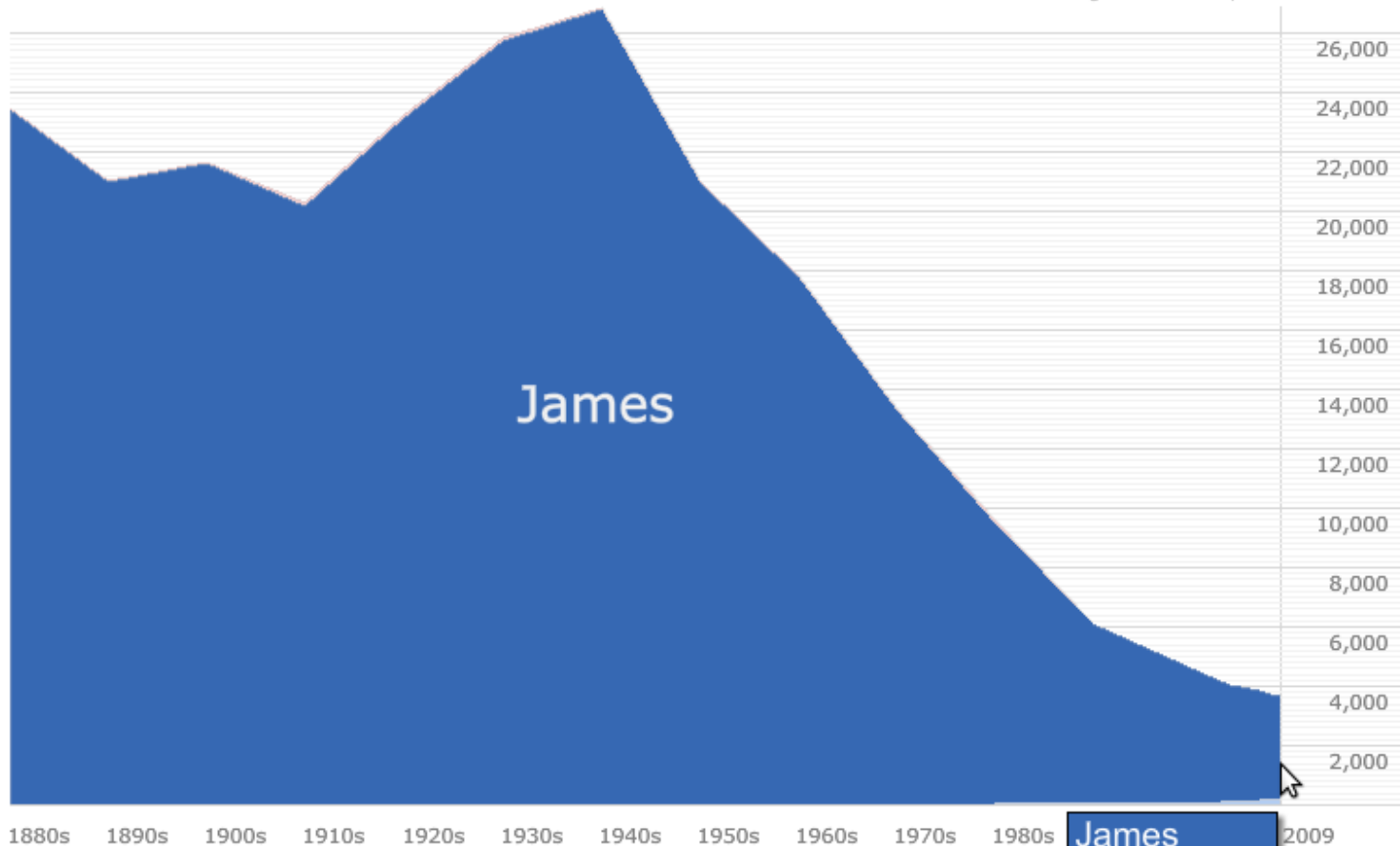
Press 'enter' to see exact matches.

☒ Both ☐ Boys ☐ Girls

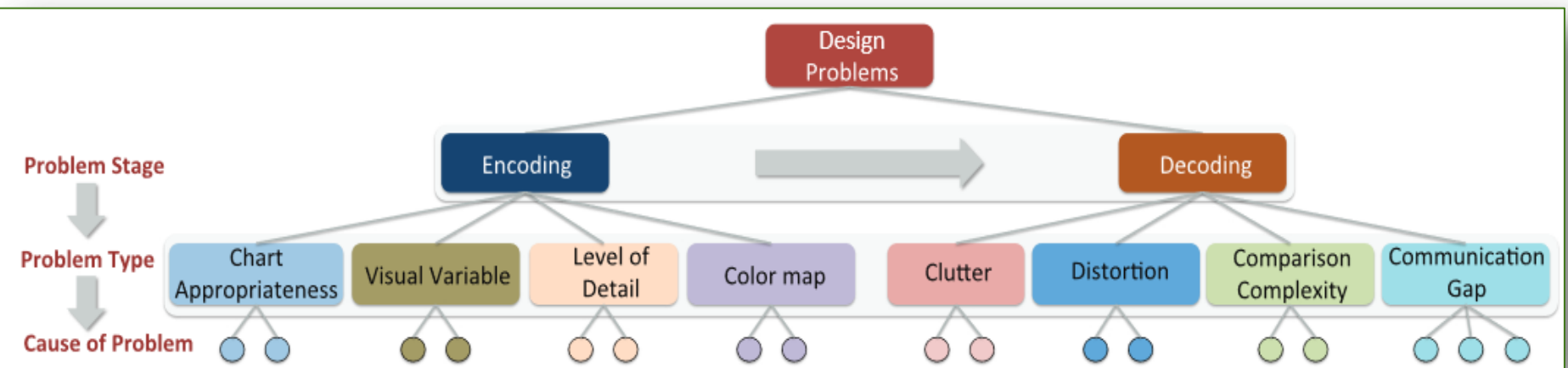
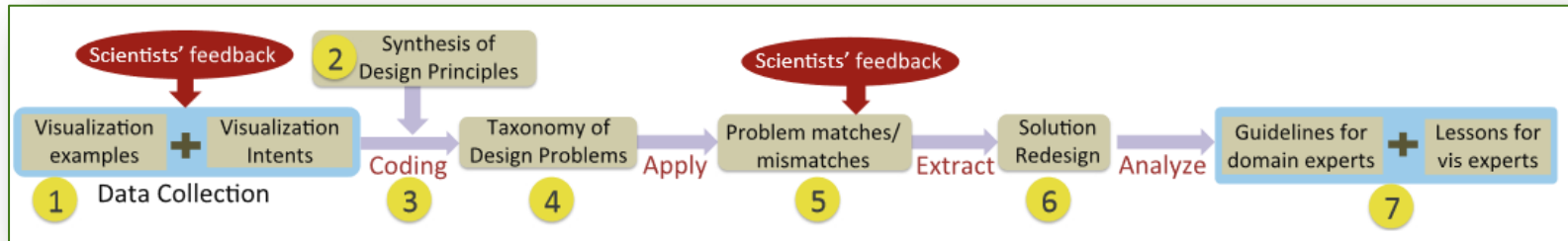
2009 rank: boys

girls

Names starting with 'JAMES' per million babies



Click a name graph to view that name. Double click to read more



Aritra Dasgupta, Jorge Poco, Yaxing Wei, Robert Cook, Enrico Bertini, Claudio T. Silva, "Bridging Theory with Practice: An Exploratory Study of Visualization Use and Design for Climate Model Comparison", *IEEE Transactions on Visualization & Computer Graphics*, vol.21, no. 9, pp. 996-1014, Sept. 2015, doi:10.1109/TVCG.2015.2413774

Workflow: Geração de conhecimento

10.1109/TVCG.2015.2467591, IEEE Transactions on Visualization and Computer Graphics

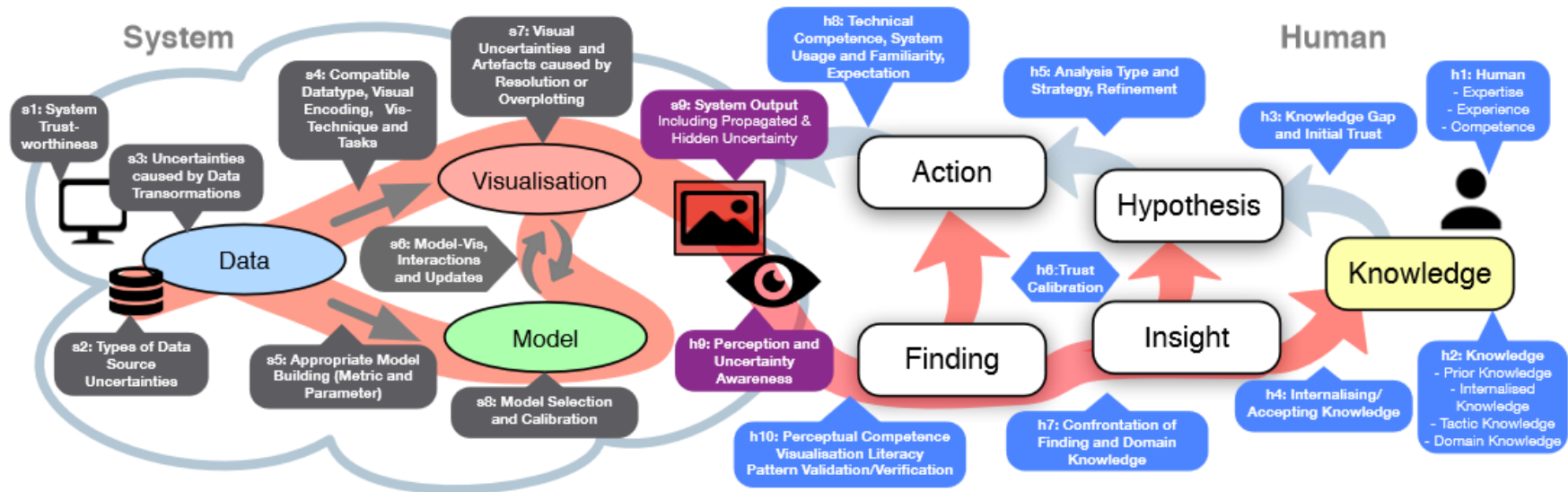


Fig. 3: Each system component may change the data and consequently introduce additional uncertainty. Human trust building within knowledge generation processes is affected by many human factors. The relation between uncertainty and trust is included as the awareness of uncertainties.

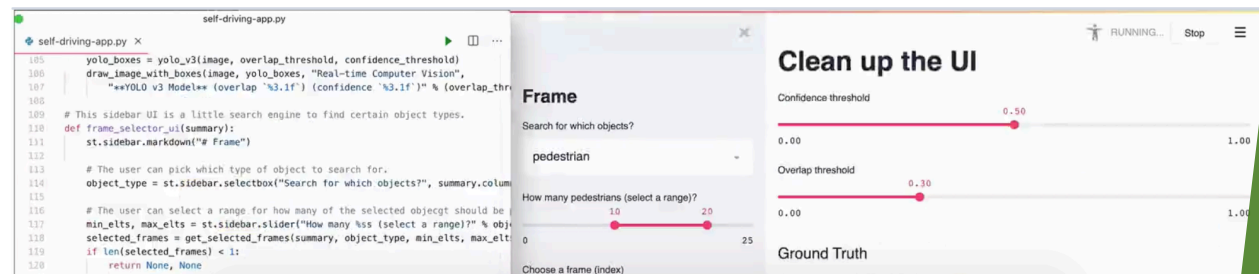
Streamlit



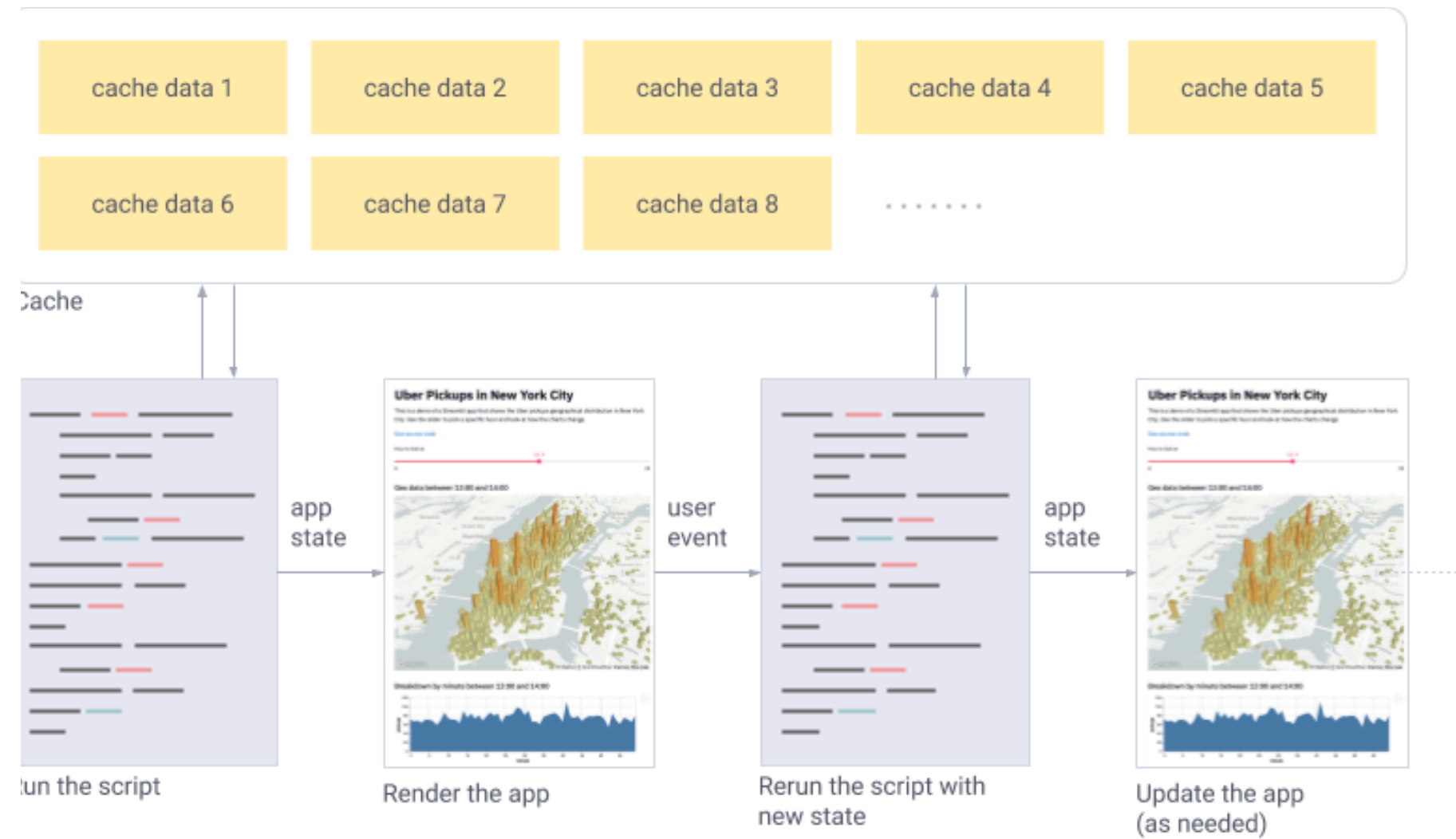
[Gallery](#) [For Teams](#) [Careers](#) [Community](#) [Docs](#) [GitHub](#) [Get Started](#)

Streamlit. The fastest way to build data apps

Streamlit's open-source app framework is the easiest way for data scientists and machine learning engineers to create beautiful, performant apps in only a few hours! All in pure Python. All for free.



Modelo de Execução



Próximos passos?

- ▶ <https://awesome-streamlit.org>
- ▶ <https://www.duckdb.org>
- ▶ <https://altair-viz.github.io>