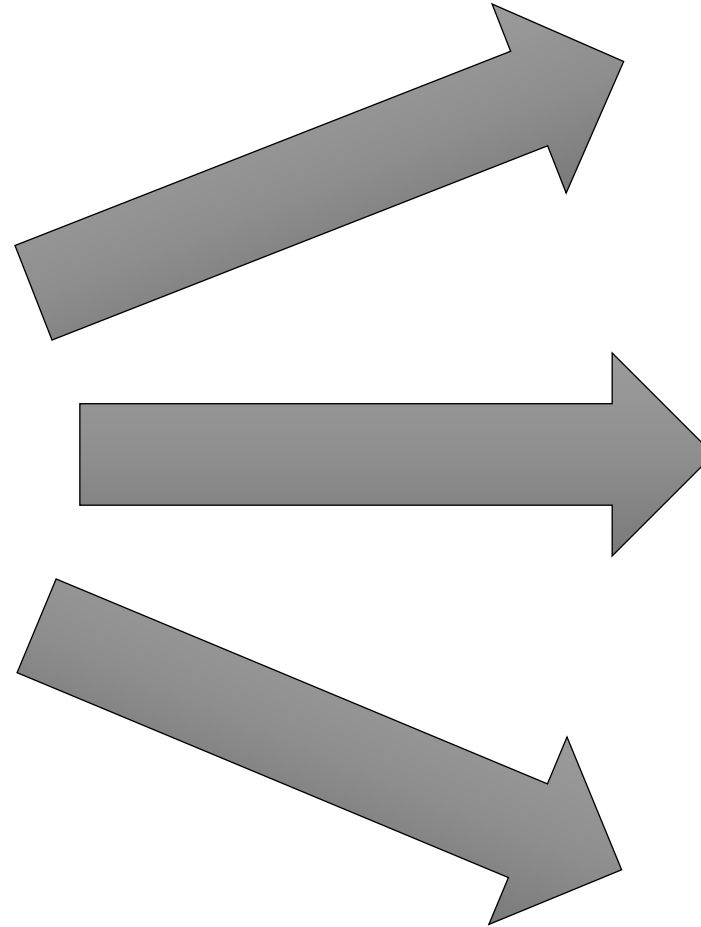


# *Próximos Passos*



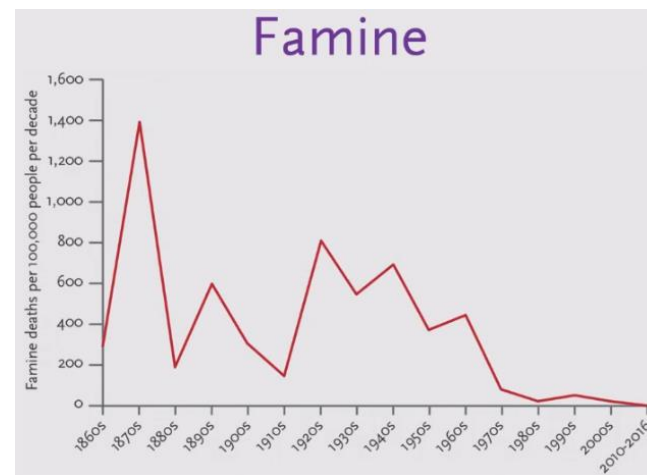
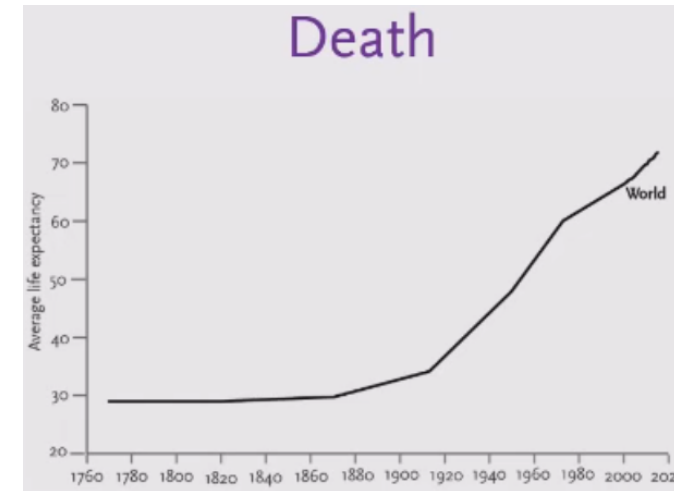
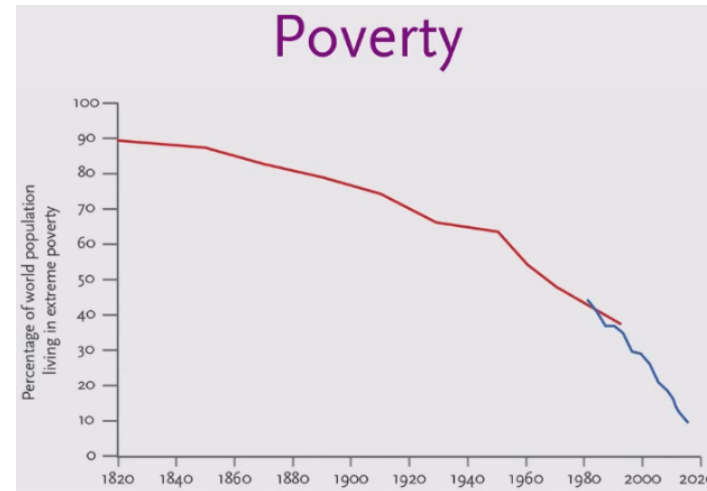
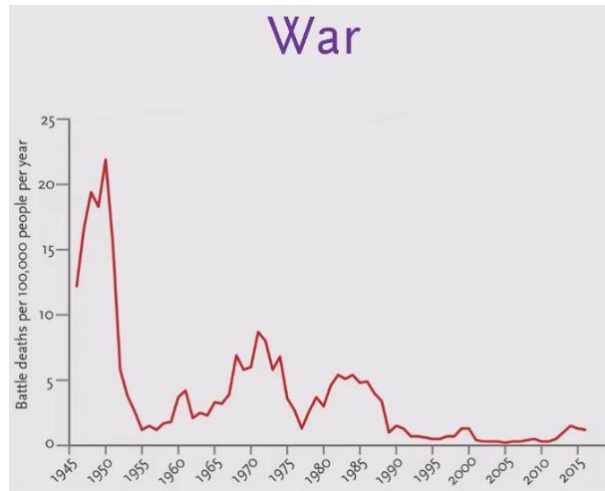
Professor: Alex Pereira

# *Próximos Passos*

- Depois de aprender a manipular dados,
  - aprender a interpretar
    - ✓ ou pelo menos não cometer erros conhecidos ao interpretá-los
- Aprender a aprender
  - Aprender condições que te levam a cometer erros
    - ✓ se preparar previamente para elas
      - Se prender ao mastro do barco, como Ulisses em Odisseia
- Inspiração na aula sobre Racionalidade do Psicólogo e Professor de Harvard
  - Steven Pinker
    - ✓ <https://harvard.hosted.panopto.com/Panopto/Pages/Sessions/List.aspx#folderID=%2255a37adc-eaae-4aa6-8a06-ab25015a4ee8%22>
      - Podcast - <https://freakonomics.com/podcast/pima-steven-pinker/>

# *Somos seres Racionais: produzimos grandes avanços nos últimos 200 anos*

- Atingimos grandes feitos
  - Viagem para a lua, foguetes reutilizáveis, veículos autônomos, implante cerebral para controlar o computador e etc



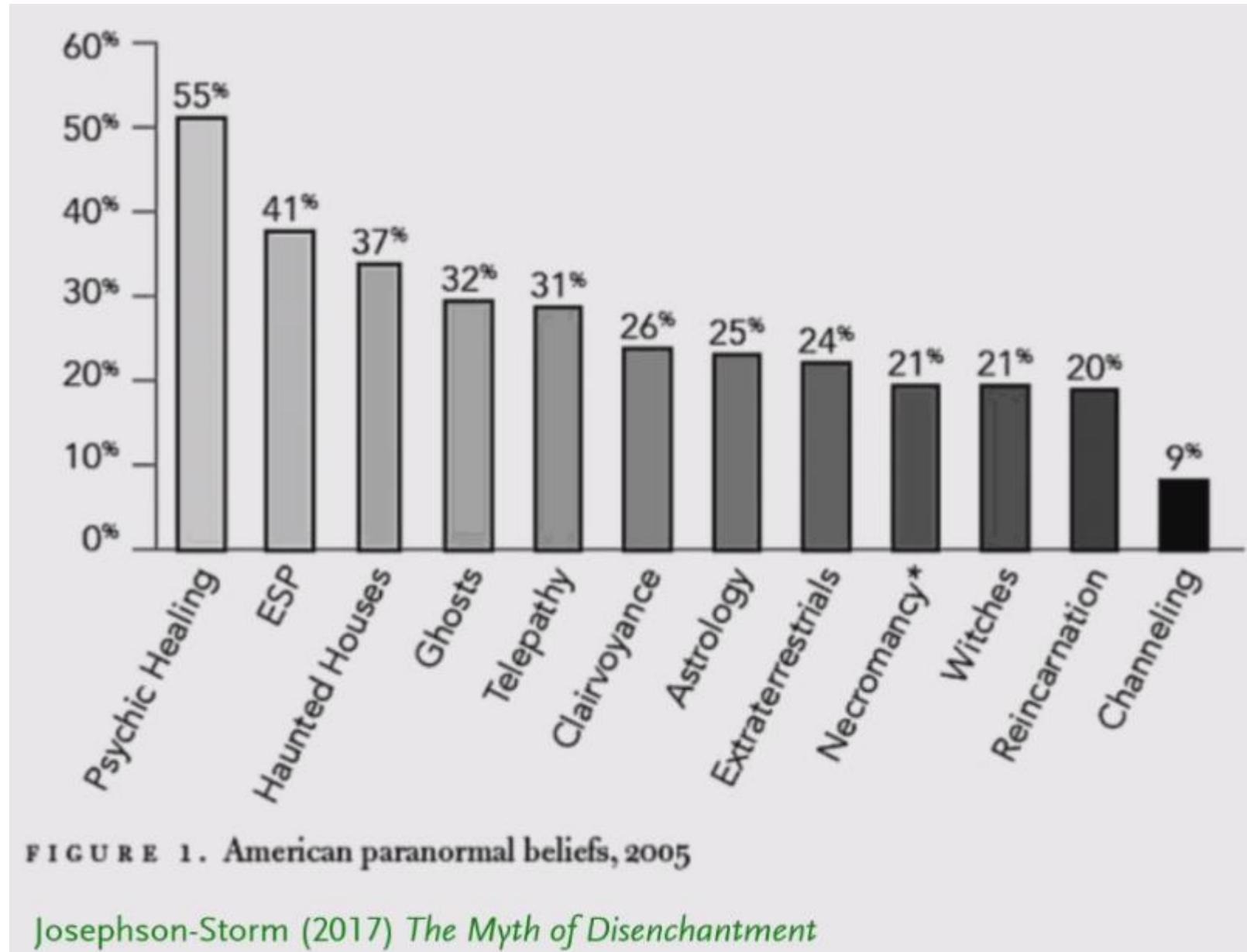


# *Paradoxalmente cometemos grandes erros em massa*



- A majority of Americans 18-24 think astrology is “very” or “sort of scientific.”

# *Paradoxalmente cometemos grandes erros em massa*





# *Há muito tempo agimos irracionalmente*



- 15<sup>th</sup>-17<sup>th</sup> C: ~60,000 witches executed for:
  - eating babies, wrecking ships, destroying crops, flying on broomsticks on the Sabbath, copulating with devils, transforming their demon lovers into cats and dogs, and making ordinary men impotent by convincing them that they had lost their penises.

## *Nosso domínio sobre a Lógica dedutiva (Se – Então)*

- Todos as cartas têm um número num lado e uma letra no outro
- Possível regra: "Se uma carta tem um **D** num lado, então ela tem um **3** no outro lado."
- Quais cartas você precisa virar para testar esta regra



# *Nosso domínio sobre a Lógica dedutiva (Se – Então)*

- Este teste foi aplicado a todo tipo de público
  - por mais de 60 anos
- A grande maioria das pessoas responde
  - "D" ou "D e 3"
- Resposta correta: "D e 7"
  - Qualquer coisa atrás do D, diferente de 3 invalida a regra.
  - Um D atrás do 7 invalida a regra.
  - Qualquer coisa atrás do 3 é irrelevante e não invalida a regra
    - ✓ Se D então 3





# *Nosso domínio sobre a Lógica dedutiva (Se – Então)*

- Erro devido ao Viés de Confirmação
  - procuramos evidências para confirmar
    - ✓ e não para falsear nossas hipóteses
  - Um bom conjunto de testes
    - ✓ testa também as possibilidades de falha
  - Conceito do Viés de Confirmação
    - ✓ é a tendência de se lembrar, interpretar ou pesquisar por informações de maneira a confirmar crenças ou hipóteses iniciais
      - Fonte: [https://pt.wikipedia.org/wiki/Vi%C3%A9s\\_de\\_confirma%C3%A7%C3%A3o](https://pt.wikipedia.org/wiki/Vi%C3%A9s_de_confirma%C3%A7%C3%A3o)

# *Lógica dedutiva – Nem sempre cometemos esse erro*

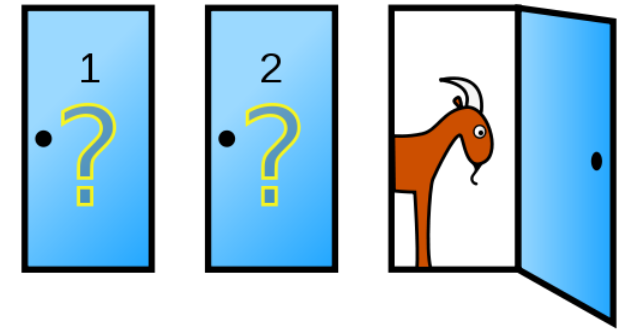
- Exemplo:

- Se um cliente de um bar está bebendo cerveja ele deve ser maior de 18 anos.
- Você é o segurança do bar e tem que assegurar o cumprimento da regra. Você checa:
  - ✓ **D** Um cara bebendo cerveja: qual sua idade? **3**
  - ✓ **F** Um cara bebendo coca-cola: qual sua idade? **Não importa**
  - ✓ **3** Um cara que é maior de 18: O que você está bebendo? **Não importa**
  - ✓ **7** Um cara que é menor de 18: O que você está bebendo? **D??**

- Equivalente a Se D então 3.

- Somos ilógicos com símbolos abstratos, e lógicos com alguns tipos de conteúdo
  - ✓ na educação formal, boa parte do aprendizado depende de conhecimento passado por meio de símbolos abstratos

# Monty Hall – Pensamento probabilístico



- Programa de TV
  - 3 portas. Uma guarda um carro. As outras duas, dois bodes;
  - O convidado do programa escolhe uma porta (Ex.: #1);
  - O apresentador abre uma das outras duas portas (não escolhidas) guardando um bode, (Ex.: #3)
    - ✓ e pergunta ao convidado: Você quer trocar de porta (#2) ou permanecer na escolhida originalmente (#1) ?
- Qual a melhor estratégia?
  - 80-90% das pessoas respondem que ficariam.
    - ✓ Alguns dizem que tanto faz.

# Monty Hall – Marilyn Vos Savant



- Ficou famosa por
  - Entrar para o livro dos recordes com o resultado do seu teste de QI e
  - Escrever uma resposta inesperada para o problema Monty Hall
    - ✓ na sua coluna da Revista Parade
      - "Você deve trocar. Na 1ª porta você tem  $1/3$  de chance de ganhar, mas na 2ª porta você tem  $2/3$  de chance de ganhar"
- 10mil cartas contestaram sua publicação
  - Uma réplica na mesma coluna da revista esquentou o debate
    - ✓ e o assunto foi parar na 1ª página do New York Times
- Por volta de 1000 cartas escritas por acadêmicos e matemáticos
  - com argumentos contrários



# Monty Hall – Marilyn Vos Savant



“You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I’ll explain. After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don’t need the world’s highest IQ propagating more. Shame!”

—Scott Smith, Ph.D., University of Florida

“I am sure you will receive many letters on this topic from high school and college students. Perhaps you should keep a few addresses for help with future columns.”

—W. Robert Smith, Ph.D., Georgia State University

“Maybe women look at math problems differently than men.”

—Don Edwards, Sunriver, Oregon

# Monty Hall – Solução



## Stay strategy:

<u>Door 1</u>	<u>Door 2</u>	<u>Door 3</u>	Result
Car	Goat	Goat	WIN
Goat	Car	Goat	LOSE
Goat	Goat	Car	LOSE

## Switch strategy:

<u>Door 1</u>	<u>Door 2</u>	<u>Door 3</u>	Result
Car	Goat	Goat	LOSE
Goat	Car	Goat	WIN
Goat	Goat	Car	WIN



# Monty Hall – Explicação Intuitiva



- Nesse exemplo o sistema 1 (intuitivo)
  - te ajuda a escolher a resposta certa.

- vos Savant's intuitive explanation:

- Say there were 100 doors. Monty opens 98 of them, revealing a goat behind all of them except #37. Should you stay with Door 1, or switch to Door 37?

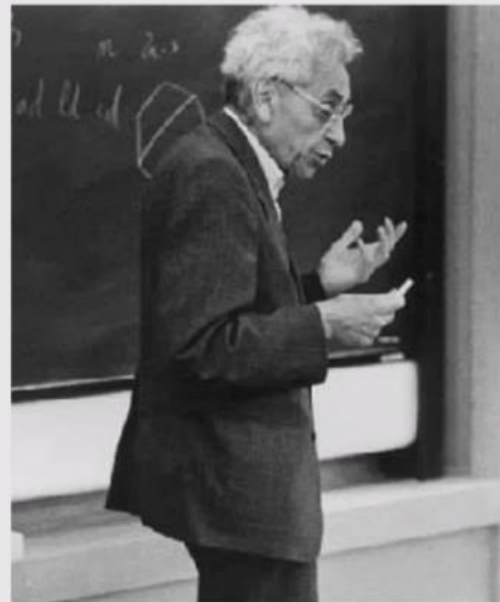
- Probabilidade expressa nossa ignorância sobre o mundo.
  - Informação adicional reduz nossa ignorância e muda as probabilidades

# Monty Hall – Paul Erdős

- Recusou-se a acreditar na solução apresentada
  - Mesmo depois de ver uma demonstração matemática e
  - Uma simulação computacional
    - ✓ executando o experimentos centenas de vezes



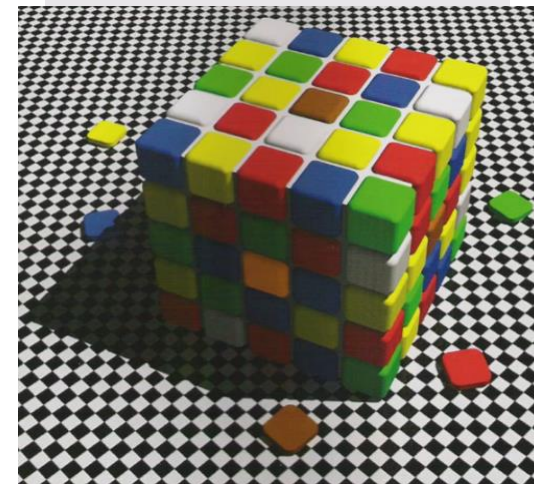
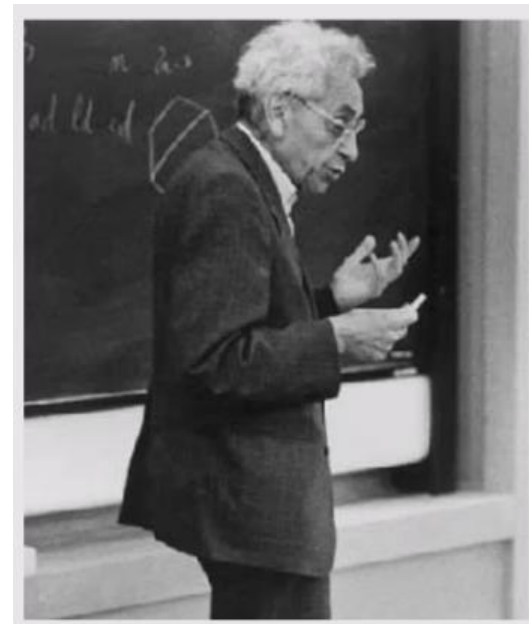
- “A renowned Hungarian mathematician, one of the most prolific of the 20<sup>th</sup> century.”



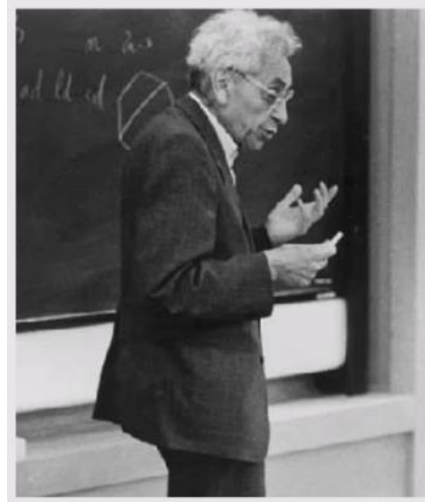


# Monty Hall – O que ocorreu no caso Paul Erdős

- Ilusão Cognitiva causa esse tipo de erro
  - Só matemáticos competentes
    - ✓ estão sujeitos a esses erros?
- Viés de Excesso de Confiança (Overconfidence Bias)
  - Nosso sentimento (subjetivo) sobre nossa certeza tem pouca relação com a probabilidade de estarmos realmente corretos
    - ✓ Entender/descrever o fenômeno da natureza vs Tentar justificar como você quer que o fenômeno ocorra.
  - Conceito:
    - ✓ leva as pessoas a sobre avaliarem suas próprias habilidades, o que provoca um comportamento de maior propensão ao risco do que o convencional para aquele indivíduo.



# Monty Hall – O que ocorreu no caso Paul Erdős



- Custo Afundado

- Conceito

- ✓ É um processo comportamental que ocorre quando colocamos dinheiro ou tempo ou qualquer outro tipo de investimento em algo,
      - e descobrimos que não conseguiremos atingir nossos objetivos.
    - ✓ Nestes casos, há uma tendência das pessoas em continuarem a investir dinheiro, tempo, etc. mesmo sabendo que isso poderá ser em vão.

- Exposição pública aumenta sua

- ✓ susceptibilidade ao custo afundado.

## *Custo Afundado*

- Eu, ao Sábado, vendo alfaces num mercado que está aberto das 8h às 13h
- Às 8h, comprei alfaces a **0,25€/kg** que fui vendendo durante a manhã.
- Às 12h59m ainda me sobram alguns kg que, se não vender, tenho que deitar ao lixo.
- Será que as devo vender a uma senhora que me oferece **0,10€/kg**?

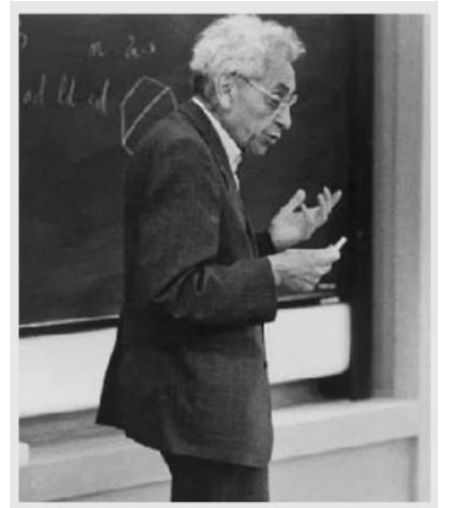
## *Custo Afundado*

- Se eu vender,  
**O benefício será de 0,10€/kg**
- Será que o custo de oportunidade são 0,25€/kg?



# *Custo Afundado*

- Se eu não vender, as alfaces vão para o lixo
- Como o custo de oportunidade quantifica o que acontece se eu não vender
- Então, **O custo de oportunidade é zero.**
- **Então, devo vender a 0,10€**
  
- O que eu paguei por elas já não conta.



Paul Erdős foi afetado pelo efeito do custo afundado?

# *Ignorando Base Rates*

Um indivíduo foi descrito por outro como segue:

- Steve é muito tímido e retraído, invariavelmente prestativo, mas com
- pouco interesse nas pessoas ou no mundo real.
- De índole dócil e organizada, tem necessidade de ordem e estrutura, e
- uma paixão pelo detalhe.
- Há maior probabilidade de Steve ser
  - um bibliotecário ou um fazendeiro?

# ***Ignorando probabilidades***

***Linda tem 31 anos, é solteira, franca e muito inteligente. É formada em filosofia. Quando era estudante, preocupava-se profundamente com questões de discriminação e justiça social, e também participava de manifestações antinucleares.***

- **Qual a alternativa mais provável?**
  - A) Linda é uma caixa de banco.
  - B) Linda é uma caixa de banco ou astronauta
  - C) Linda é uma caixa de banco e é ativa no movimento feminista.

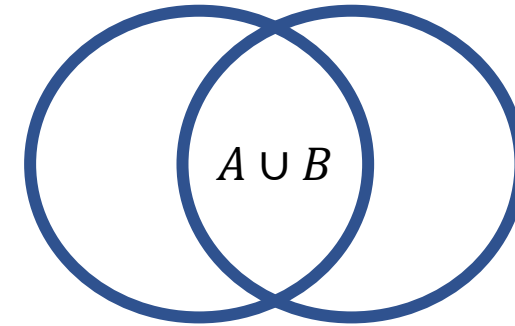
# Análise dos Problemas apresentados anteriormente

- Steve

- Ignoramos os base rates (as taxas de recorrência a priori)
  - ✓ Existem muito mais fazendeiros do que bibliotecários
- Respondemos intuitivamente outra pergunta
  - ✓ Qual descrição melhor se encaixa no perfil de Steve  $\neq$  Qual a probabilidade

- Linda

- Teoria de conjuntos
  - ✓  $P(A \cup B) \geq P(A) \geq P(A \cap B)$
- Respondemos intuitivamente outra pergunta
  - ✓ Qual descrição melhor se encaixa no perfil de Linda



- O enunciado te desviou o foco da parte importante do problema?
- Sob uma pergunta difícil para a qual não conhecemos a resposta
  - Mas conhecemos a resposta de outra pergunta mais simples
    - ✓ Intuitivamente, damos como resposta esta outra



# *Heurística da Disponibilidade / Viés da Disponibilidade*

- Conceito
  - um fenômeno no qual as pessoas predizem a frequência de um evento,
    - ✓ ou a proporção de uma população,
  - baseando-se no quão fácil conseguem lembrar de um exemplo.
- Num estudo científico, pesquisadores pediram aos participantes
  - Considerar causas de morte aos pares
    - ✓ diabetes e asma, ou derrame e acidentes.
  - Para cada par, as pessoas indicavam a causa mais frequente e
    - ✓ estimavam a proporção das duas frequências

# *Heurística da Disponibilidade / Viés da Disponibilidade*

- Resultado da Pesquisa de percepção
  - Derrames causam quase o dobro de mortes de todos os acidentes combinados, mas
    - ✓ 80% dos participantes avaliaram a morte acidental como mais provável.
  - Tornados foram vistos como assassinos mais frequentes do que asma, embora
    - ✓ esta última provoque vinte vezes mais mortes.
  - Morte por acidentes foi avaliada como mais de trezentas vezes mais provável do que morte por diabetes,
    - ✓ mas a proporção verdadeira é 1:4.
- Porque há maior disponibilidade (lembrança)
  - de alguns tipos de informações?
    - ✓ Por causa da frequência da exposição a estas informações

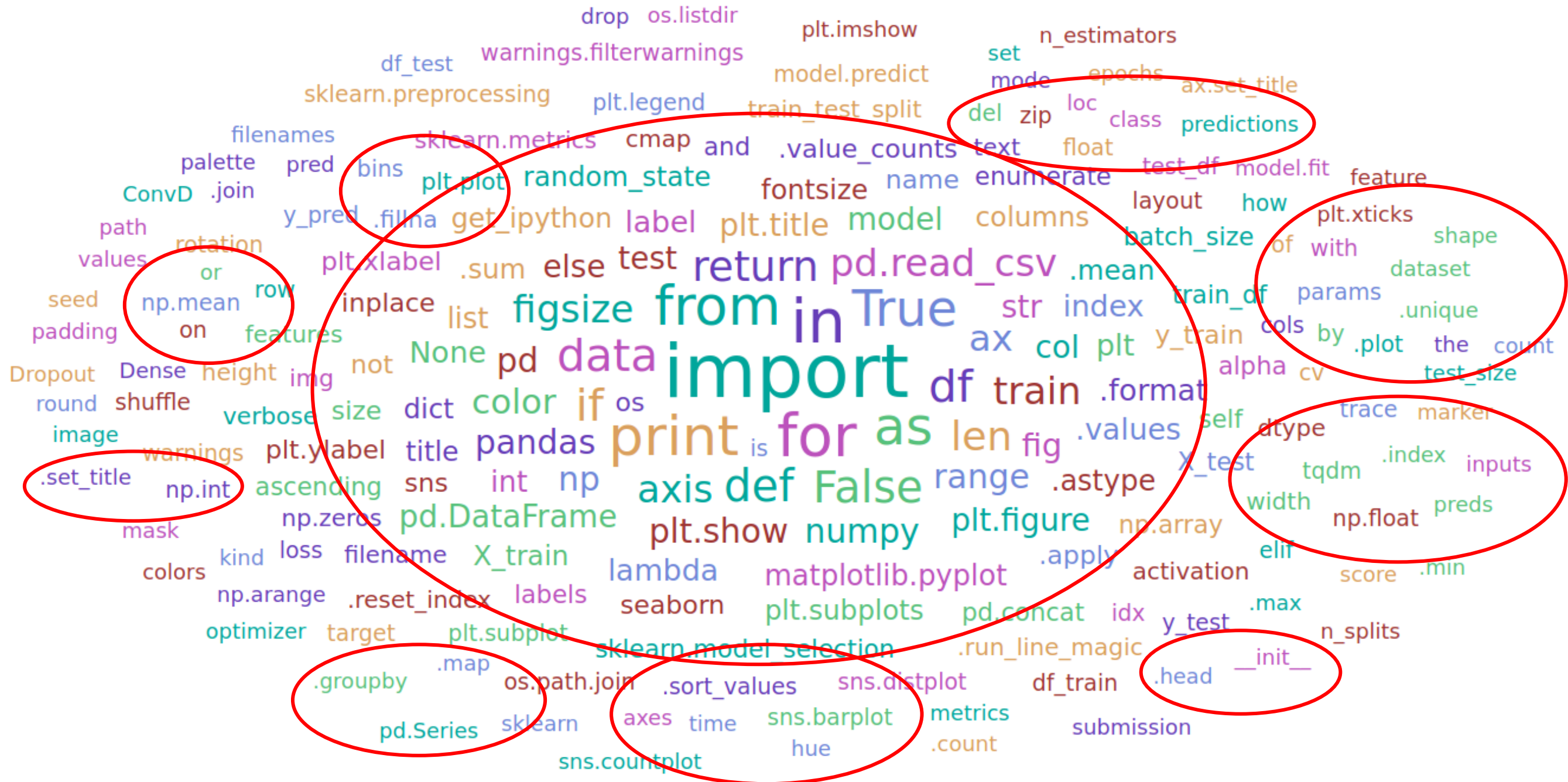
# *Notícias e Heurística de Disponibilidade*

- Jornalismo apresenta uma amostra sistematicamente enviesada (negativa)
  - da experiência/contexto mundial
    - ✓ Daniel Kahneman, Richard Thaler, Nassim Taleb e Steven Pinker
- As coisas "desmoronam" rapidamente (recebem mais destaque)
  - mas melhoram lentamente (recebem menos destaque)
- Cobertura de acontecimentos repentinos e a
  - Orientação comercial (gerar audiência)
    - ✓ Contribuem para esse viés
  - Exemplo: Headlines sobre acompanhamento do Covid no G1
    - ✓ A métrica mais negativa é colocada no título (headline)
      - Quando a média móvel diminui, ela não aparece no título

# *Notícias e Heurística de Disponibilidade*

- Analise de Sentimento de 1 milhão de Headlines de Notícias (Austrália)
  - Na média, o sentimento é negativo durante todo o mês
  - ✓ <https://www.kaggle.com/richel145/analysis-of-a-million-news-headlines>
- Fake News têm 70% mais chance de ser compartilhado
  - Nos EUA, 126mil rumores se espalharam por ~3 milhões de pessoas
  - ✓ <https://science.sciencemag.org/content/359/6380/1146>
- Chats em Grupo nos Aplicativos de Mensagem
  - Qualidade editorial?
  - Evitar o padrão (consumo aleatório e oportunístico de informação)
  - ✓ Dar preferência para consumir a informação pelo autor, e por manifestação ativa de vontade
    - A disponibilidade das informações na sua memória forma a sua opinião.
    - Coloque informação de qualidade pra dentro.

# Roadmap do Vocabulário





# *Como progredir mais rápido: Desenvolver hábitos para melhorar o desempenho*

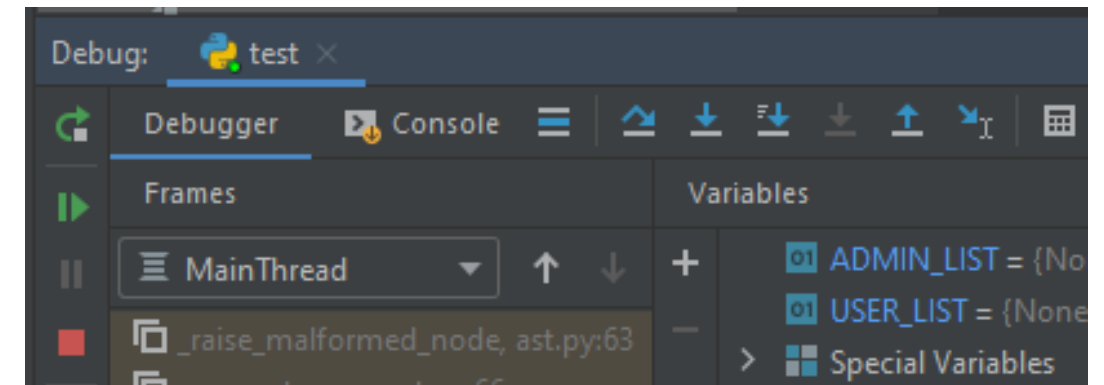
- Meta-análise: melhorar o seu processo individual de trabalho
  - Refletir sobre o que te levou a cometer cada erro
    - ✓ e como evita-lo da próxima vez (Ciclo de PDCA)
  - Exemplo 1:
    - ✓ Perceber que ao desenvolver um código diretamente numa função
      - você está aumentando o tempo do ciclo de feedback
        - Note que isso está te atrasando e corrija seu hábito
  - Exemplo 2:
    - ✓ Perceber que você comete menos erros se abrir e fechar cada símbolo de escopo conjuntamente
      - `"//div[@class='button']"`

# *Como progredir mais rápido: Desenvolver hábitos para melhorar o desempenho*

- Encontre oportunidades para encurtar os tempo de receber feedbacks

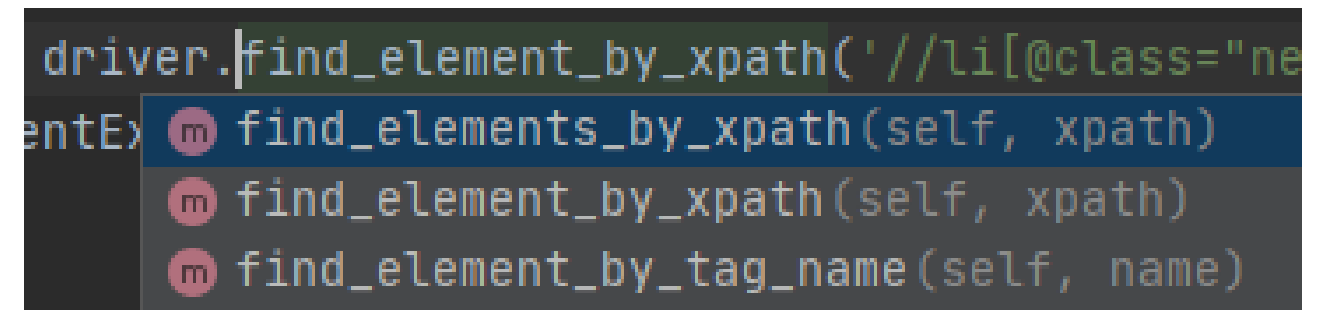
- Desenvolvimento interativo com Debug

- ✓ break point e watch expression.



- Autocomplete de nomes de funções

- ✓ Economiza por volta de 50% da digitação



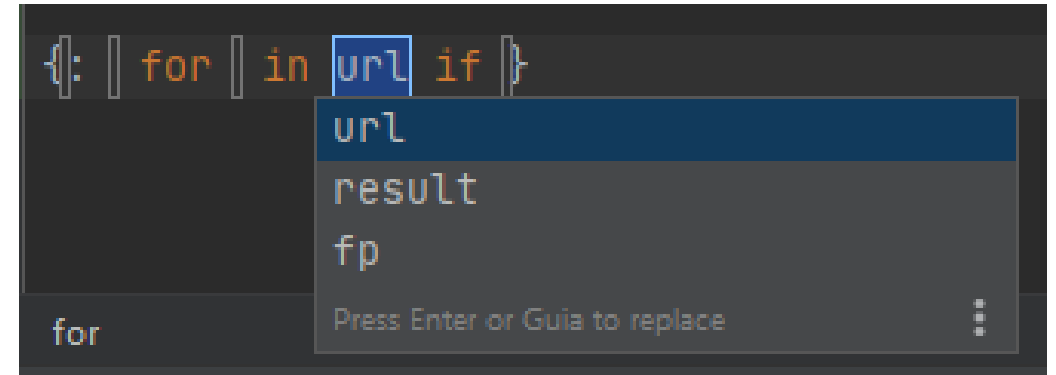
# *Como progredir mais rápido: Desenvolver hábitos para melhorar o desempenho*

- Encontre oportunidades para encurtar os tempo de receber feedbacks

- Live templates

- ✓ Exemplo: Dict comprehension

- Você pode criar seus templates

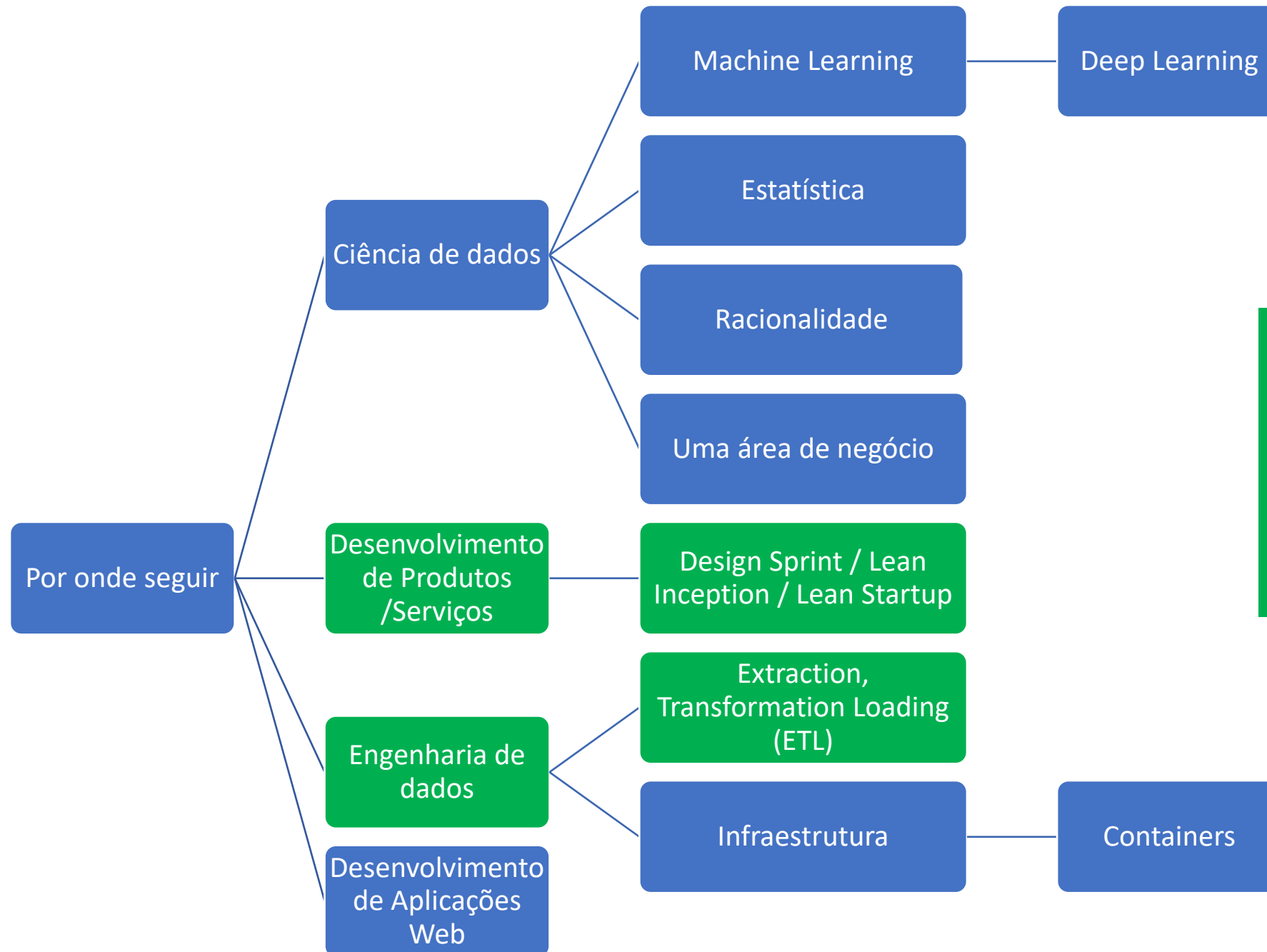


- Leia a mensagem de erro / log !!

- é o caminho mais curto entre você e a solução do seu problema

- ✓ Qualquer coisa diferente disso não é racional.

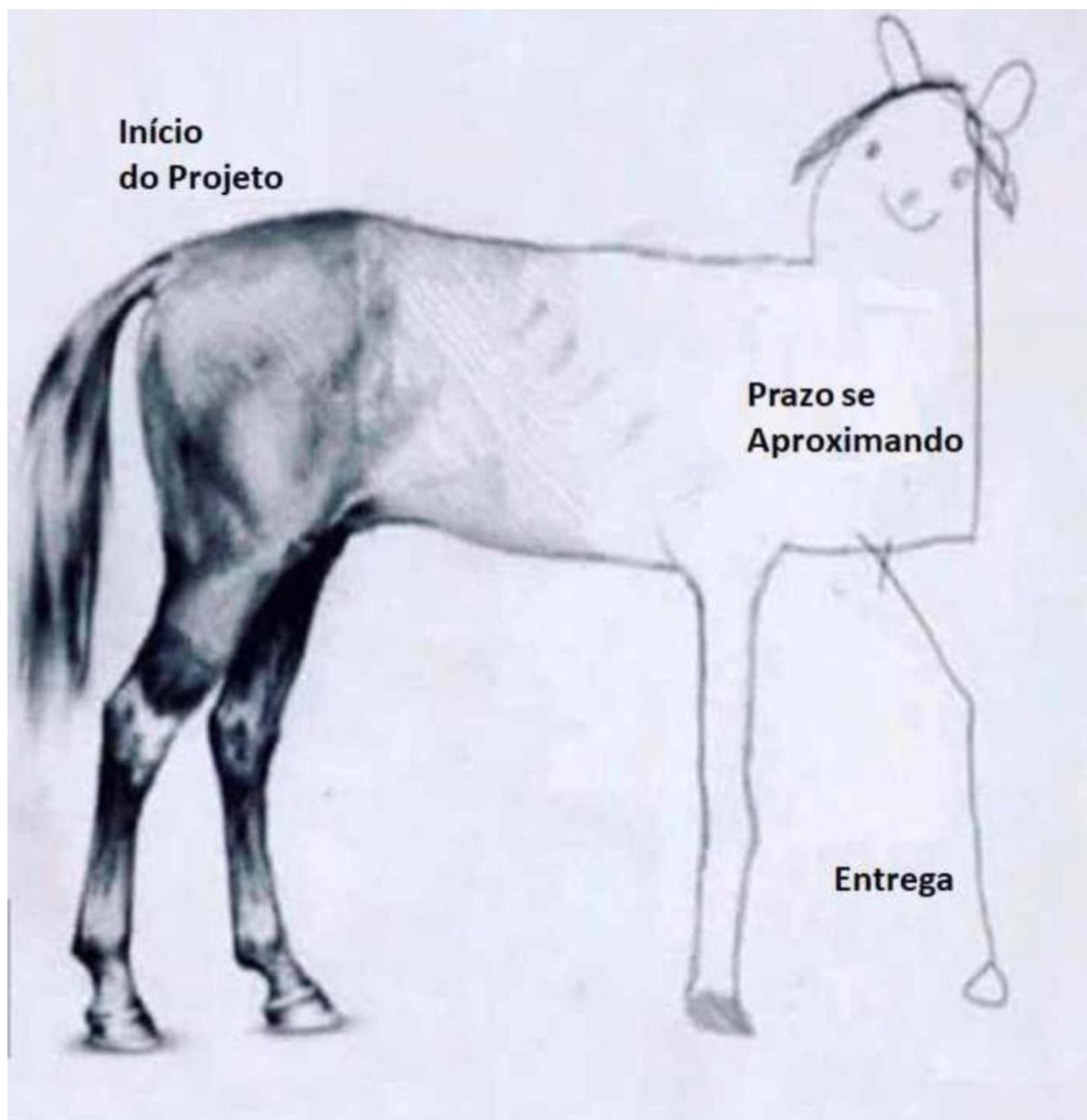
# *Agregue o máximo possível de habilidades*



Estas são atividades elementares, não fazê-las bem feitas implica em grandes prejuízos.

***Se você falha ao planejar, está planejando falhar***

*Benjamin Franklin*

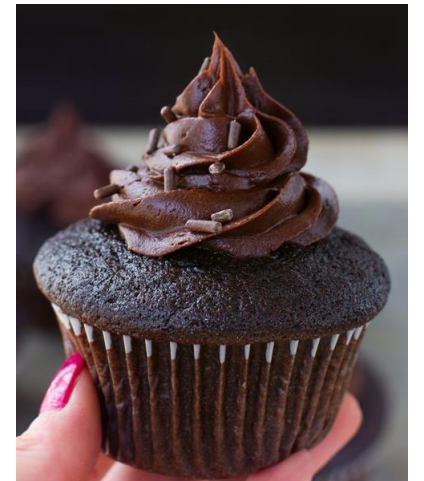




# Conceituação de Agilidade

- Sinônimo de
  - Flexibilidade
    - ✓ Celeridade e baixo custo de mudanças
  - Proatividade
    - ✓ A equipe identifica problemas e corrige
  - Rapidez
    - ✓ Entrega contínua
    - ✓ Cupcake

Pequenos Lotes  
(Small batches)



# Agilidade vs Pressa

- Pressa

- Tem conotação pejorativa (imprudência/negligência)
  - ✓ Exemplo: pular etapas essenciais.
    - Para cumprir um pseudo deadline.
- Fazer bem feito leva o mesmo tempo
  - ✓ de fazer mal feito.
    - Se tem que ser feito, tem que ser bem feito.
    - Se não for fazer bem feito, não faça.



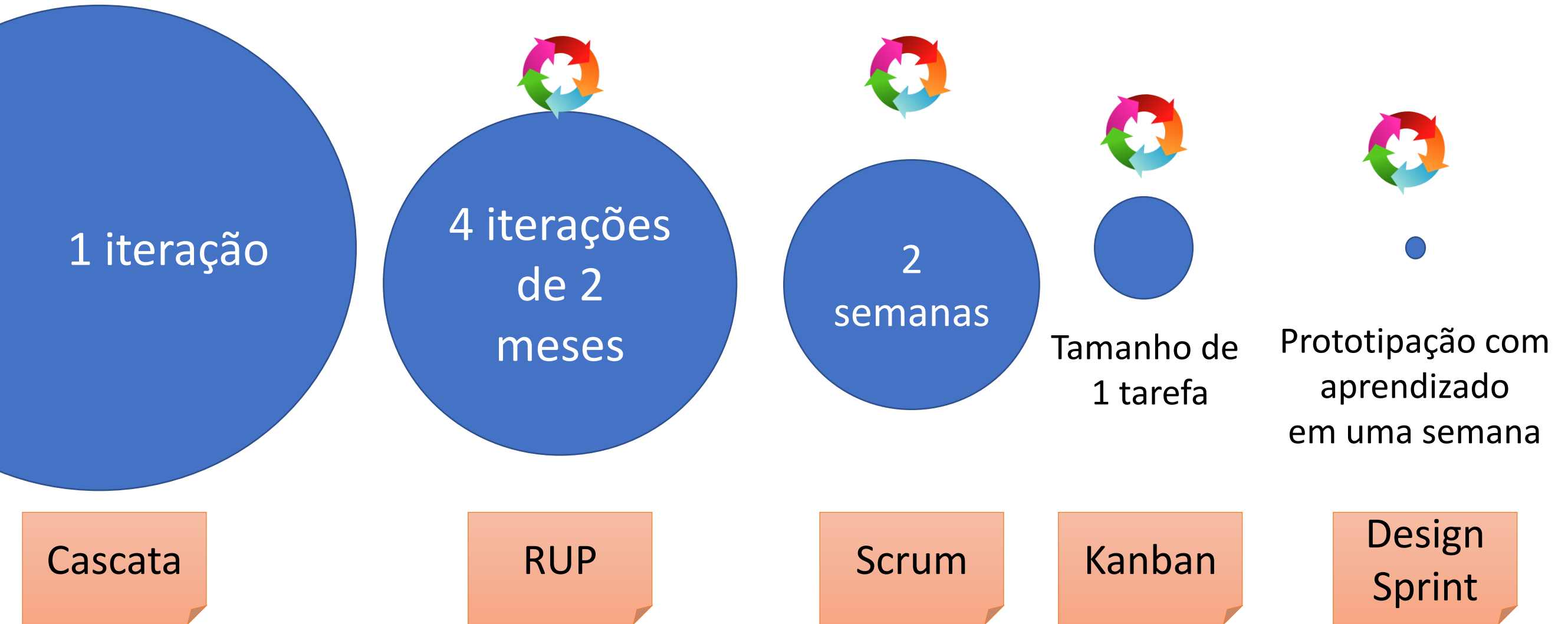
Fonte: Alex Pereira

- Agilidade

- Maximizar a quantidade de trabalho não realizado
  - ✓ Atingindo o mesmo objetivo
    - Fazer um cupcake bem feito



# *Tamanho decrescente das iterações nas metodologias de gestão*



- Feedback antecipado e frequente

# *Valores do Manifesto Ágil*

- Como ler os valores: Priorizar isso mais do que aquilo.
- Indivíduos e interações
  - Mais do que processos e ferramentas;
- Software em funcionamento mais
  - do que documentação abrangente;
- Colaboração com o cliente
  - Mais do que negociação de contratos;
- Responder a mudanças
  - Mais do que seguir um plano

# *Valores do Manifesto Ágil depois do Lean Startup*

- Visão de time e disciplina
  - mais do que indivíduos e interações
    - ✓ mais do que processos e ferramentas
- Aprendizado Validado
  - mais do que software funcionando
    - ✓ mais do que documentação abrangente
- Descoberta/entendimento do cliente
  - mais do que colaboração com o cliente
    - ✓ mais do que negociação de contratos
- Iniciar mudanças
  - mais do que responder a mudanças
    - ✓ mais do que seguir um plano

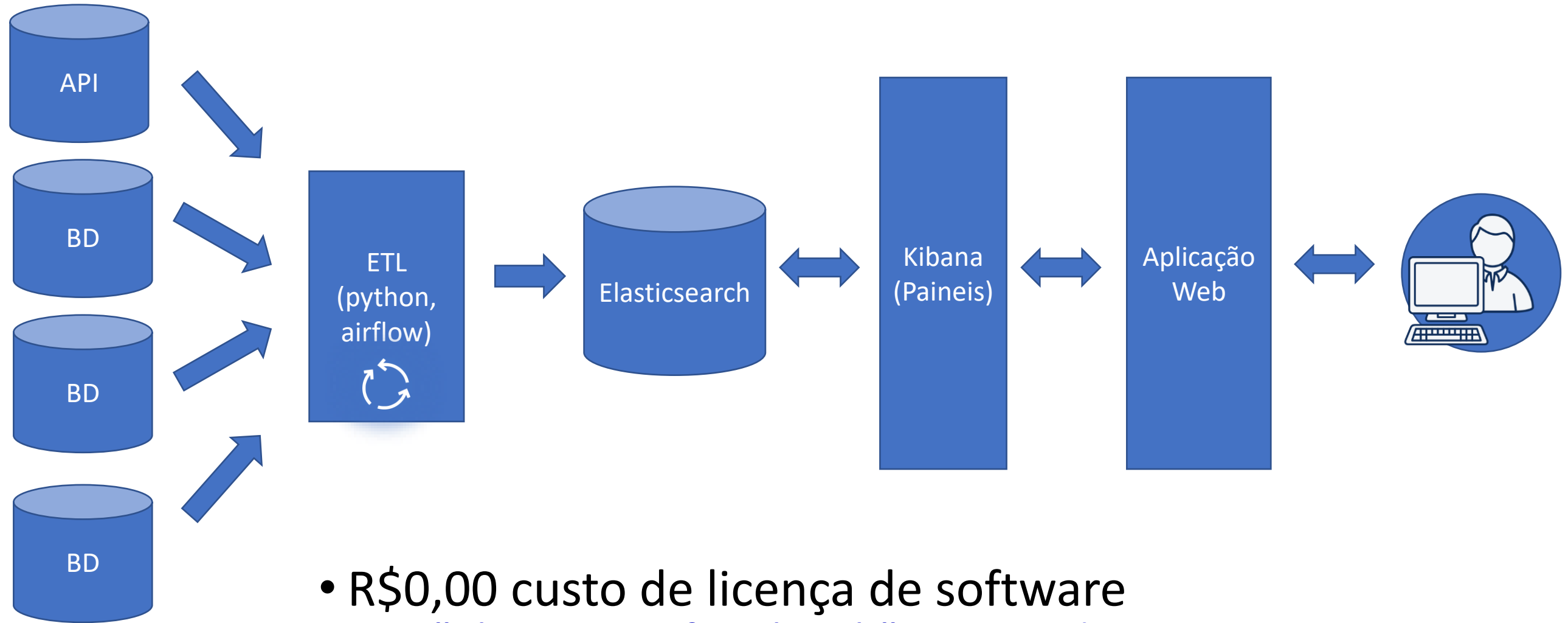


# Recomendação de Ferramentas

- Qual a melhor ferramenta depende do seu contexto
  - Competência prévia, recursos \$ disponíveis.

Atividade	Ferramentas
Machine Learning	Scikit-Learn
Deep Learning	Fastai
Estatística	Statsmodel
ETL	Python, Airflow
Base de Dados	Elasticsearch
Painéis	Kibana, Superset, Metabase
Desenvolvimento de Aplicações Web	Flask
Desenvolvimento de Produtos	Design Sprint e Lean Inception
Containers	Docker e Kubernetes

# Arquitetura de Solução de Paineis com Elasticsearch



- R\$0,00 custo de licença de software
  - "There is no free lunch": Demanda investimento em capacitação
    - Cientista de Dados, Engenheiro de Dados, Desenvolvedor Web e Especialista em Infraestrutura de TI

# Arquitetura de Solução com Elasticsearch

Secretaria de Governo Digital (SGD)  
Departamento de Experiência do Usuário de Serviços Públicos



Área do Gestor de Serviços

Página Inicial > Estatísticas de Acesso Web

orgao\_nome: Ministério da Economia (ME) × + Add filter

Período

⌚ Quick

Último 1 ano

Filtro

Órgão

Ministério da Economia (ME) ×

Serviço

Select...

Taxa de rejeição média

0,647

Taxa de Rejeição Média

Número de visualizações

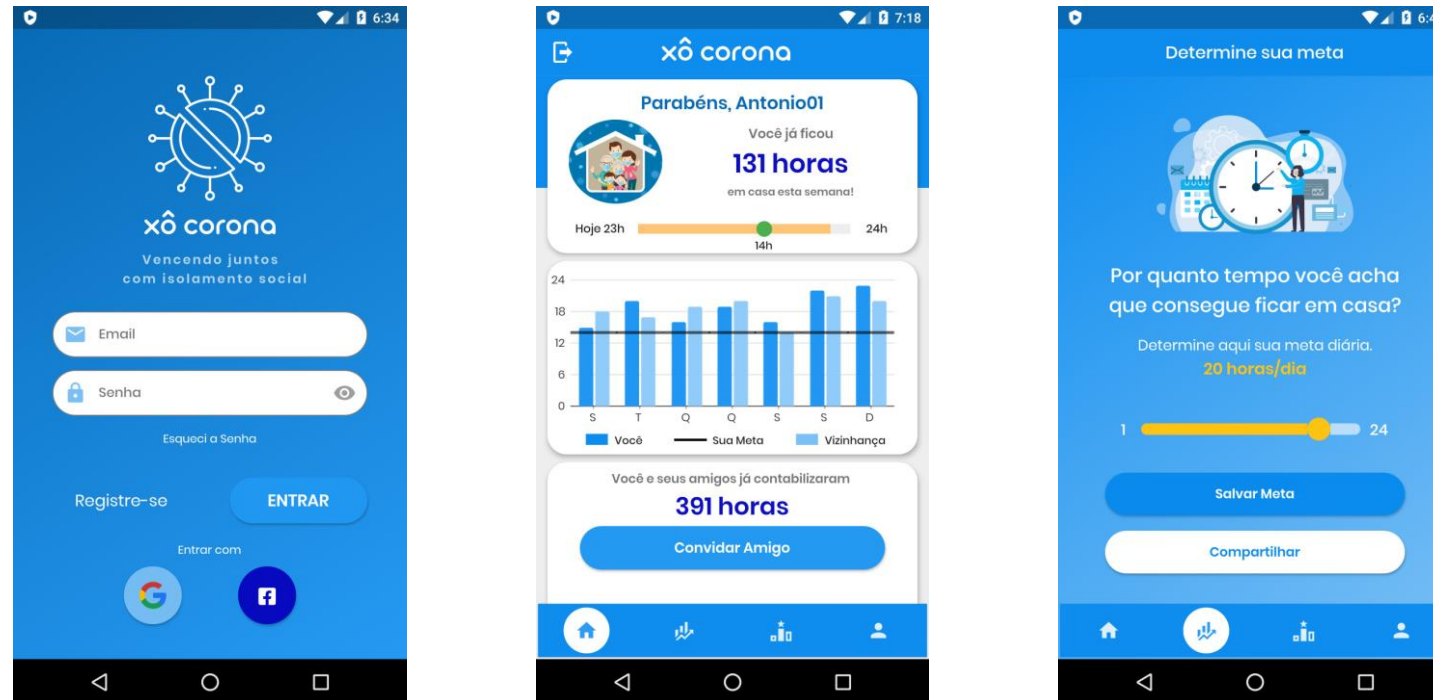
27.060.969

Visualizações

Tabela de serviços

Serviços	Visualizações	Taxa de Rejeição
Sacar o Abono Salarial	11.642.905	0,8
Solicitar o Seguro-Desemprego	7.738.494	0,666
Obter a Carteira de Trabalho	4.297.268	0,671
Cadastrar recurso relativo ao seguro-desemprego	816.319	0,502
Buscar emprego no Sistema Nacional	235.670	0,684

# Backend de Aplicação com Elasticsearch



```
s = Search(using=config.es, index=config.INDEXES[2]['index']) \
    .filter('term', last_monday=lm.strftime("%Y-%m-%d"))
s = s[:0]
s.aggs.bucket('per_percentil', 'percentiles', field='totalhometime', percents=[10, 99])
response = s.execute()
percentis = response.aggregations.per_percentil.to_dict()
```

# *Trabalho Final*

1. Estabeleça um objetivo de análise de dados.
2. Escolha uma ou mais bases de dados do seu interesse.
3. Faça análises estatísticas sobre esta base de dados e apresente argumentos, achados ou conclusões não triviais baseados nos parâmetros calculados por você.
4. O formato deve ser um Jupyter notebook contando uma história com as seguintes seções: Apresentação da Base de Dados, Motivação, Investigação (Com código fonte), e Principais achados ou conclusões.
5. Trabalhos utilizando uma base de dados entre 2 mil e 100 mil linhas, valem no máximo 8. Entre 100 mil e 500 mil linhas, 9. E mais do que 500 mil, 10.
6. A base de dados deve ser pública ou acessível ao Professor.
7. Prazo de Entrega: **07/09/2020** até as **23h59** no horário de Brasília. Atrasos terão **descontos exponenciais a cada dia**.

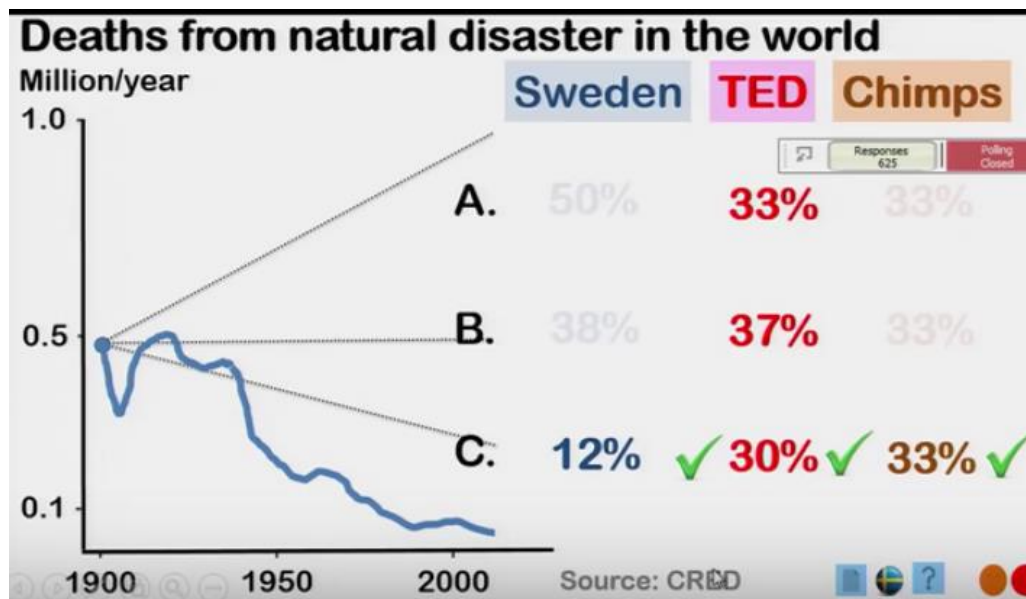


# *Entrega dos Exercícios*

- Os exercícios poderão ser submetidos na plataforma do Jupyter hub até o dia **13/09/2020** às 23h59 no horário de Brasília.
- Depois desta data o ambiente do Jupyter Hub será reiniciado
  - Sem dados de exercícios submetidos.
  - ✓ **Salve suas soluções antes desta data!**
- O ambiente do classroom continuará disponível
  - E a solução dos exercícios será disponibilizada lá
- Nota Final
  - $(\text{Nota Exercicios} * 0.7 + \text{Nota Trabalho Final} * 0.3)$
- Nota de Aprovação
  - 70%

# *Regras de bolso para escapar de erros elementares de julgamento*

- Nossa intuição sobre a evolução de indicadores de
  - pobreza, desigualdade de gênero, mortes por catástrofes naturais e outras estatísticas corriqueiras,
- Costuma ser pior do que o desempenho
  - de chimpanzés sobre o mesmo problema
    - ✓ pior do que o desempenho aleatório (jogar uma moeda)



# *Regras de bolso para escapar de erros elementares de julgamento*

- Usando a intuição (Rules of thumb)
  - As coisas estão melhorando ao longo do tempo
    - ✓ E não piorando (misconception 1)
  - Prefira assumir que os eventos são descritos por uma curva monomodal
    - ✓ do que uma curva bi-modal (misconception 2)
  - Desenvolvimento social ocorreu antes dos países se tornarem ricos
    - ✓ E não o contrário. (misconception 3)
  - Assuma que você vai exagerar a sua percepção de risco
    - ✓ Sobre assuntos emotivos (Terrorismo, catástrofes naturais, acidentes aéreos, ataques de tubarão e etc)
      - Todos esse matam muito menos que causas corriqueiras (acidentes de trânsito, por exemplo)
- Tente não depender dessas heurísticas
  - Sempre que possível, consulte uma fonte oficial.