https://stochasticsolutions.com/pdf/toronto2023-errors-of-interpretation.pdf

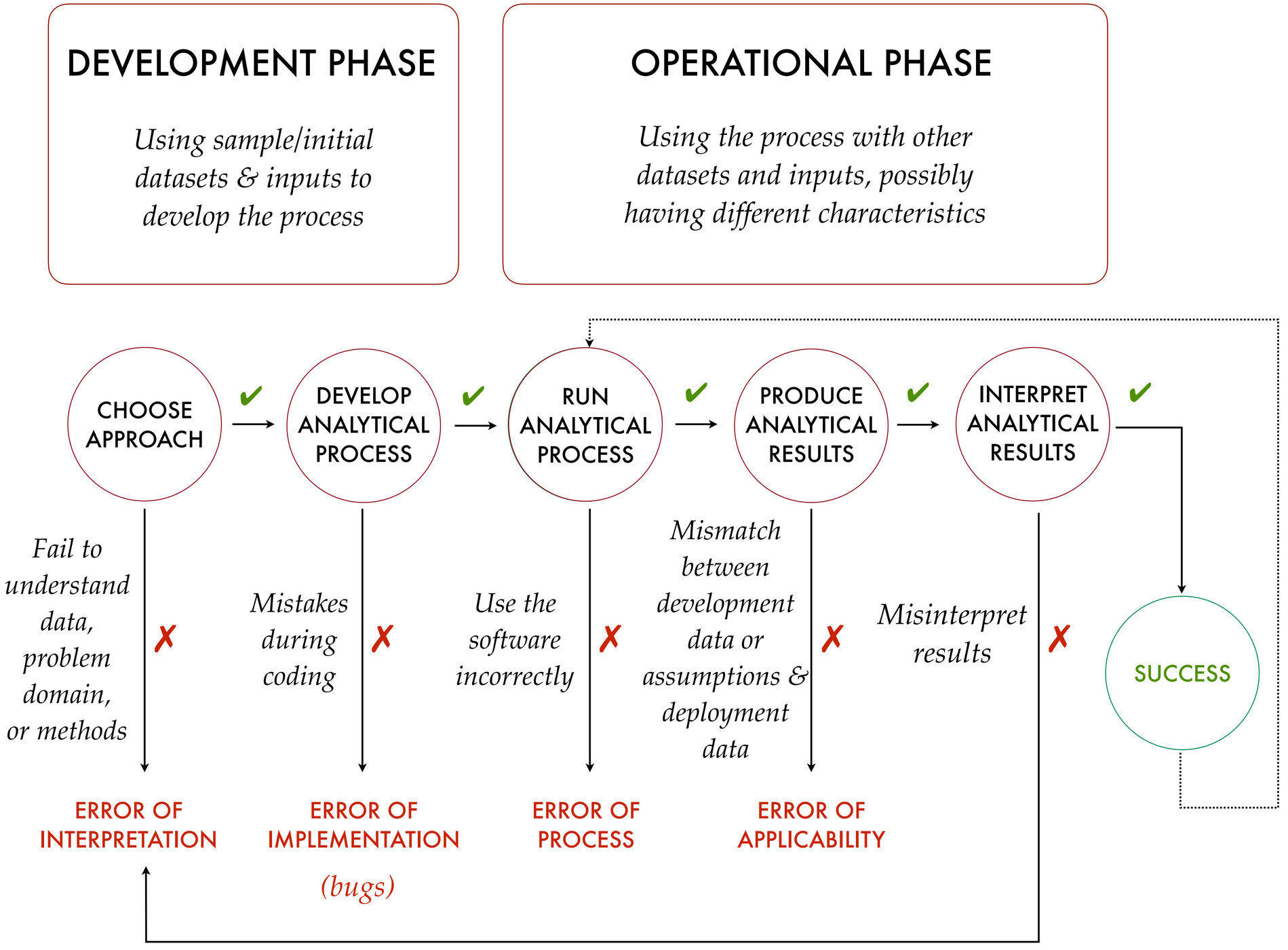
# ERRORS OF INTERPRETATION (a.k.a. TYPE VI ERRORS)

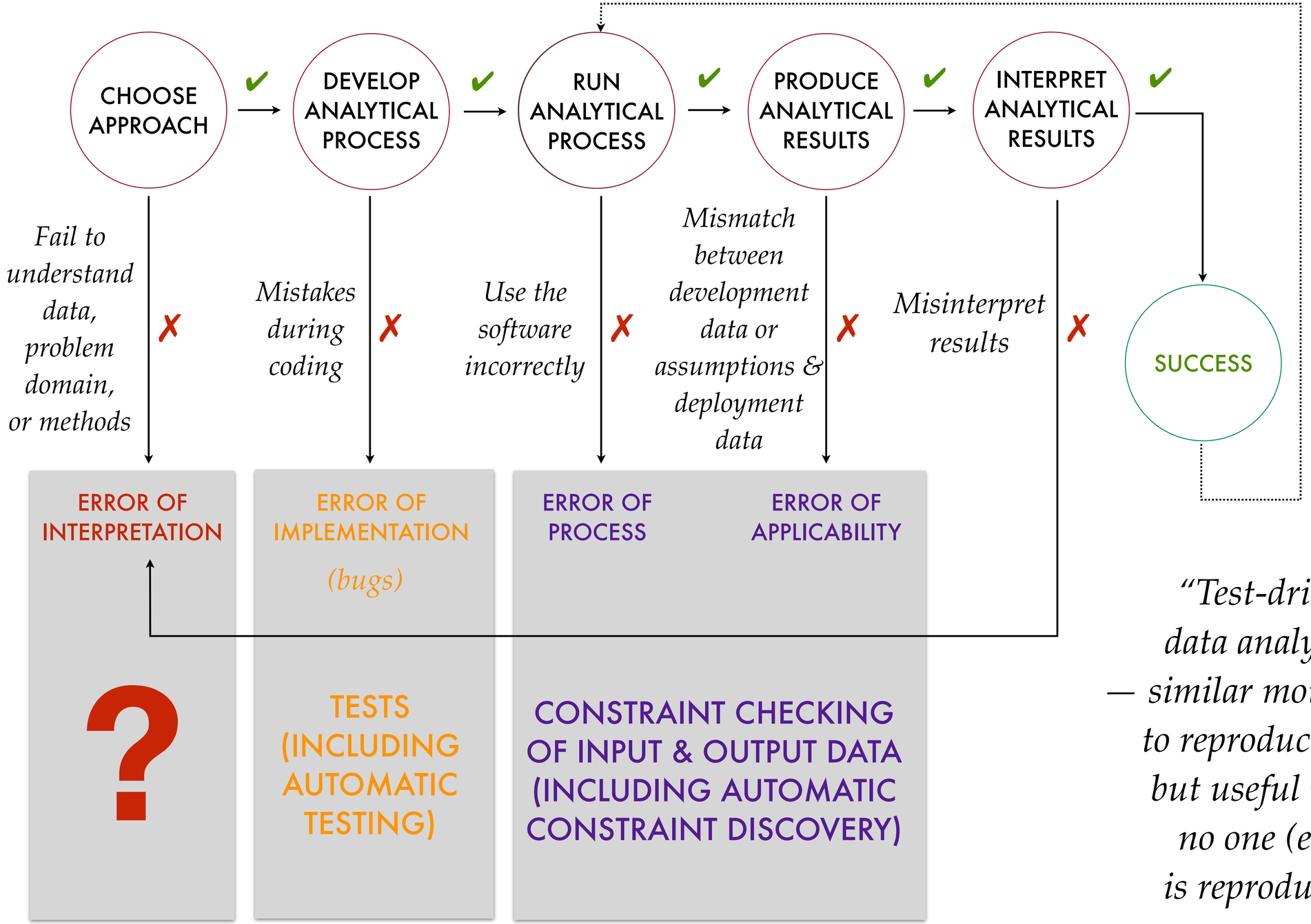
**Nicholas J. Radcliffe**
Stochastic Solutions Limited
& Department of Mathematics, University of Edinburgh
& Smart Data Foundry

2023 Toronto Conference on Reproducibility


Stochastic Solutions

SMART DATA FOUNDRY


DEVELOPMENT PHASE
Using sample/initial datasets & inputs to develop the process
OPERATIONAL PHASE
Using the process with other datasets and inputs, possibly having different characteristics
CHOOSE APPROACH
DEVELOP ANALYTICAL PROCESS
RUN ANALYTICAL PROCESS
PRODUCE ANALYTICAL RESULTS
INTERPRET ANALYTICAL RESULTS
SUCCESS
Fail to understand data, problem domain, or methods
Mistakes during coding
Use the software incorrectly
Mismatch between development data or assumptions & deployment data
Misinterpret results
ERROR OF INTERPRETATION
ERROR OF IMPLEMENTATION
(bugs)
ERROR OF PROCESS
ERROR OF APPLICABILITY




TDDA APPROACH

*"Test-driven data analysis" — similar motivations to reproducibility but useful when no one (else) is reproducing*

*First do no harm*

# Mars Climate Orbiter

**NASA (SI)**
Newton-seconds

v.

**Lockheed Martin (FPS)**
Pounds (force)
-seconds

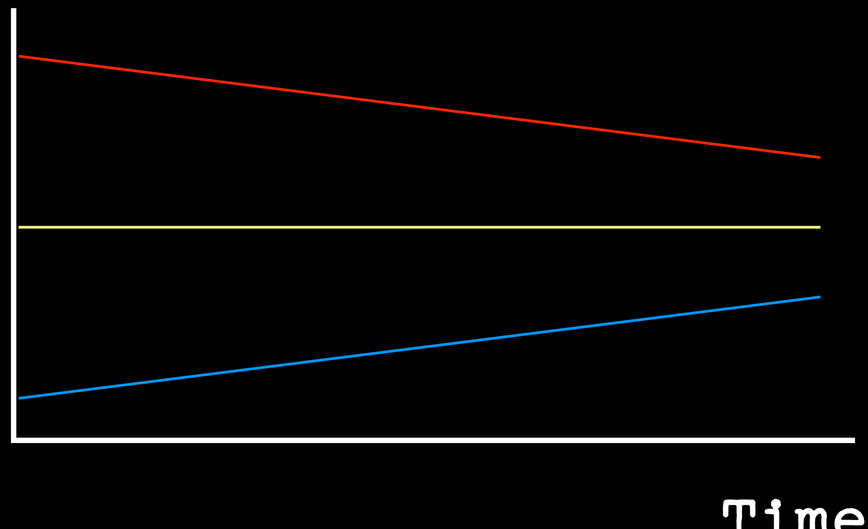
| | |
|---|---|
| m | metres |
| m | miles |
| m | milli |
| m | million |
| M | Million (Mega) |
| M | Thousand |
| Mi | $2^{20}$ (1,048,576) |
| MM | Million |
| k | Thousand |
| K | $2^{10}$ (1024) |
| K | Kelvin |

| | |
|---|---|
| bn | $10^{9}$ |
| bn | $10^{12}$ |
| B | $10^{9}$; $10^{12}$ |
| 90° | π/2 |
| 90° | 45% alcohol |
| 90° | nearly boiling (°C) |
| 90° | wear suscreen (°F) |
| calories · Calories · kcal | |
| pt | 20 fl oz · 16oz |
| pt | 1/72.27” · 1/72” |

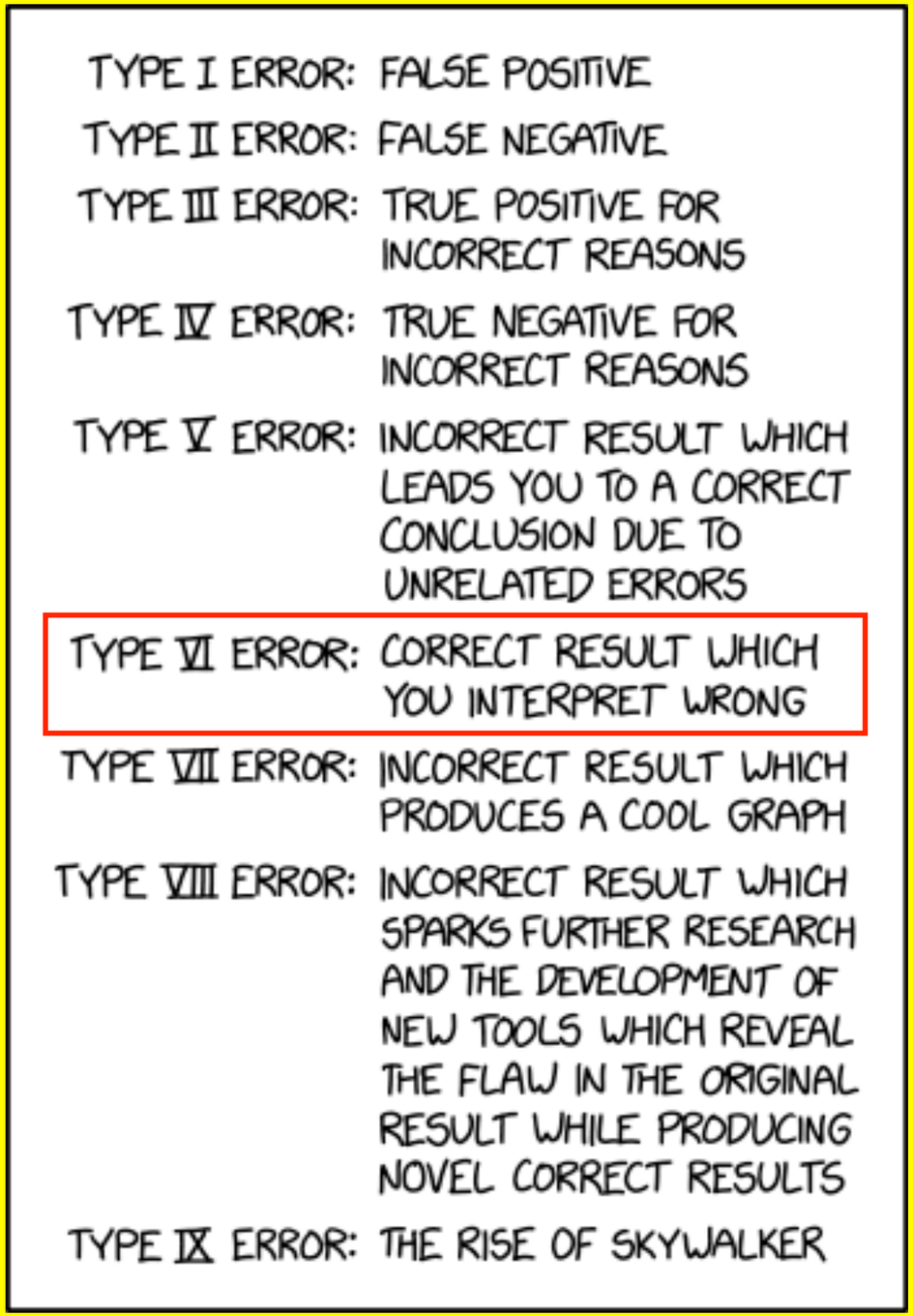
Which class are we predicting?

99.9983%

# Regression to the mean



# “Type I” & “Type II” Errors




https://xkcd.com/2303

01/02/12

# Significant Figures & Spurious Precision

# Table 2: World Water

| | $km^3$ | Per cent |
|---|---|---|
| **Fresh Water** | | |
| Clouds | 20,000 | 0 |
| Continental Water | 9,000,000 | 1 |
| Ice | 30,000,000 | 2 |
| **Salt Water** | | |
| Oceans | 1,300,000,000 | 97 |
| **Total water** | 1,339,020,000 | 100 |

Source: **Not:** Sustainability: A Systems Approach. A M H Clayton & N J Radcliffe

# Water

**Table 2:** World water

| | $km^3$ | *Per cent* |
|---|---|---|
| *Fresh water* | | |
| Clouds | 20,000 | 0 |
| Continental water | 9,000,000 | 1 |
| Ice | 30,000,000 | 2 |
| *Salt water* | | |
| Oceans | 1,300,000,000 | 97 |
| Total water | c 1,300,000,000 | 100 |

Source: Sustainability: A Systems Approach. A M H Clayton & N J Radcliffe

# Percentage Changes

## “Relative vs. Absolute risk”

| | Increase |
|---|---|
| 1% → 1.1% | +10%<br>+0.1pp |

# JUNK CHARTS

Dual Axis

Non-uniform scale

False zero*

False zero colour*

Area, Volume

Inverted

Bezos charts

Unclear labels

Unclear tick labels

No units

Questionable lines of best fit

* When zero is meaningful

# GRAPHING BEST PRACTICES

Annotate

Maximize Data Ink

Minimize chart junk

Direct labelling

Error bars

Pie charts are OK!

Units

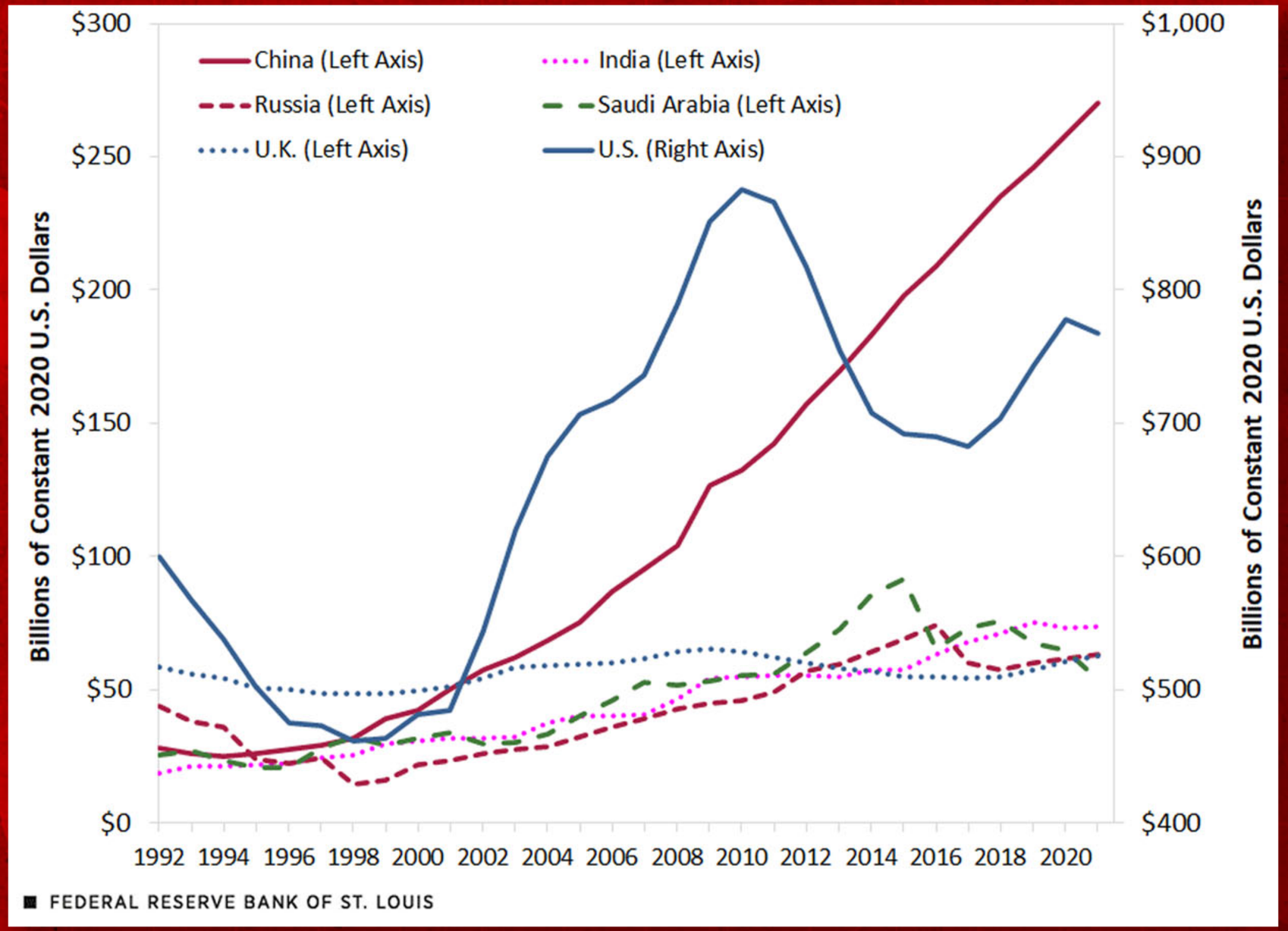
Zoomed sections for detail & context

Broken axes where required

```
>>> datetime.date(2101,12,2).strftime('%y/%d/%m')
'01/02/12'
```

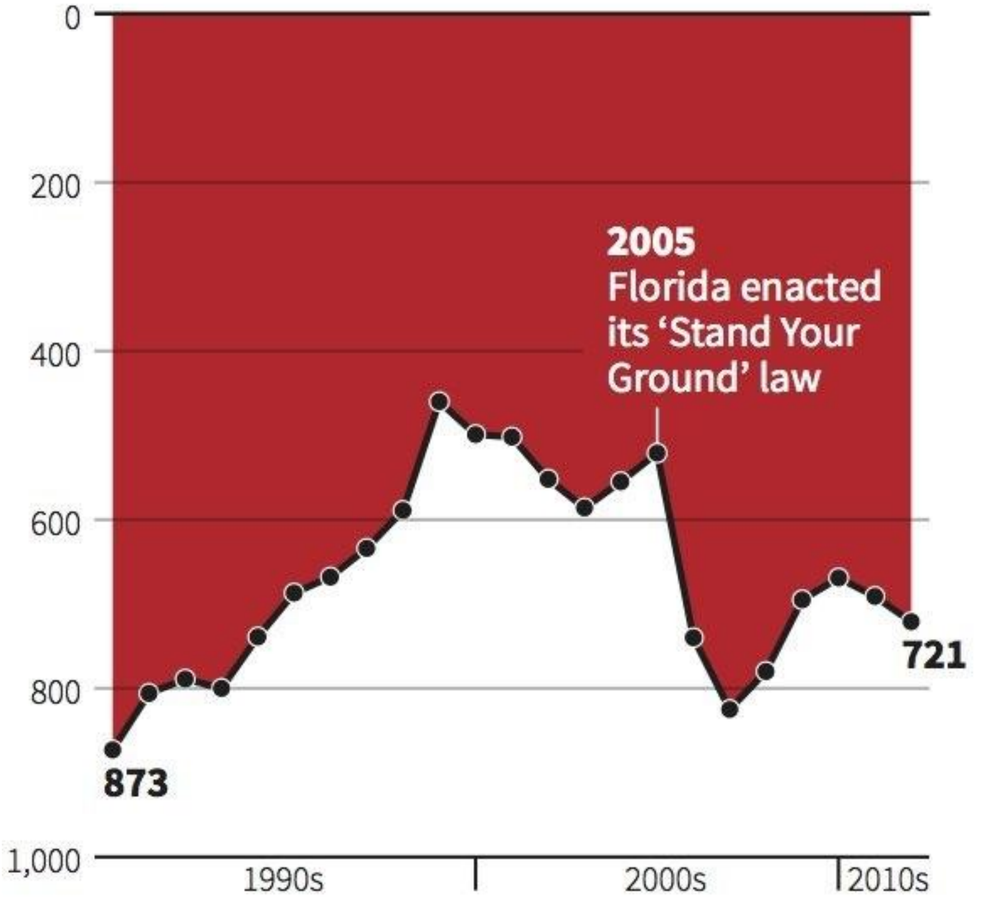
# 01/02/12

(2 Dec 2101)


China (Left Axis)
India (Left Axis)
Russia (Left Axis)
Saudi Arabia (Left Axis)
U.K. (Left Axis)
U.S. (Right Axis)
$300
$250
$200
$150
$100
$50
$0
$1,000
$900
$800
$700
$600
$500
$400
Billions of Constant 2020 U.S. Dollars
Billions of Constant 2020 U.S. Dollars
1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020
FEDERAL RESERVE BANK OF ST. LOUIS

# Gun deaths in Florida

Number of murders committed using firearms




Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

https://stochasticsolutions.com

http://tdda.info

https://github.com/tdda

njr@StochasticSolutions.com

https://linkedin.com/in/njradcliffe

#tdda

@njr@zirk.us

https://stochasticsolutions.com/pdf/toronto2023-errors-of-interpretation.pdf