

Predicting Elections

A Brief Introduction to Data Science

Nuwan I. Senaratna

3 Mini Case Studies

1. How to **Fact-Check** Election Predictions
2. How to **Predict** Elections
3. How to Predict Elections (in **Real Time**)

CASE 1

How to **Fact-Check** Election Predictions



මහ මැතිවරණය පිළිබඳව සුජීත් නිශාන්තයන්ගේ ආනාවකි | Astrological Predictions On 2020 Election

Wishwa Danuma 74K views · 3 months ago

In this video Astrologer, Sujith Nishantha Yogi reveals who will win the 2020 Sri Lanka Parliamentary Election from his astrological viewpoint. Subscribe for more videos:



22 Election Districts Summary | Sri Lanka Parliament Election 2020 Final Seats Predictions | 2020

Sahir Tamil Poetry24 · 4.8K views · 3 months ago

22 Election Districts Summary | Sri Lanka Parliament Election 2020 Final Seats Predictions | 2020 | National List Predictions #SrilankaParliamentElection2020 #ParliamentElection2020SrilankaPredictions



Sri Lanka election prediction-2020

maha samatha · 4K views · 3 months ago

Sri Lanka election prediction-2020 පාර්ලිමේන්තු මැතිවරණයෙන් දිනන්නෙ කවුද? හියවර ජයග්‍රහණයා ගැන තක්කෙටම අනාවැකි කී අපගේ මහා සම්මත YouTube channel යෙහි මෙවරත් අනාවැකි. ජනාධිපති වරණය පිළිබඳව අනාවැකිය-Link- ...



Why does Sri Lanka's parliamentary election matter? | Inside Story

Al Jazeera English 52K views · 2 months ago

They're known for their role in ending the 26-year war against Tamil separatists. Now, they are seeking to extend their family's rule over the island nation. President Gotabaya's party aims to win a two-thirds

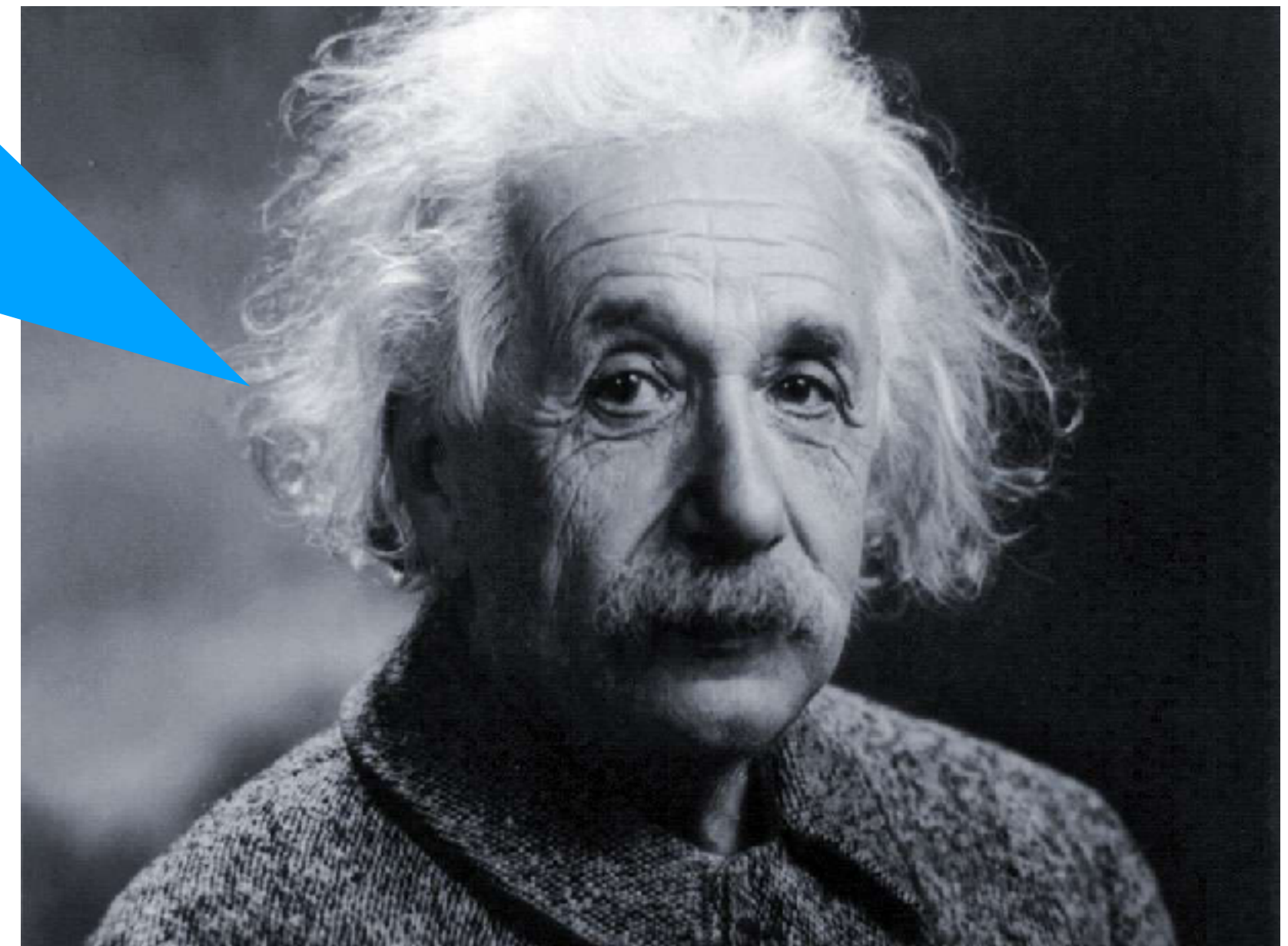


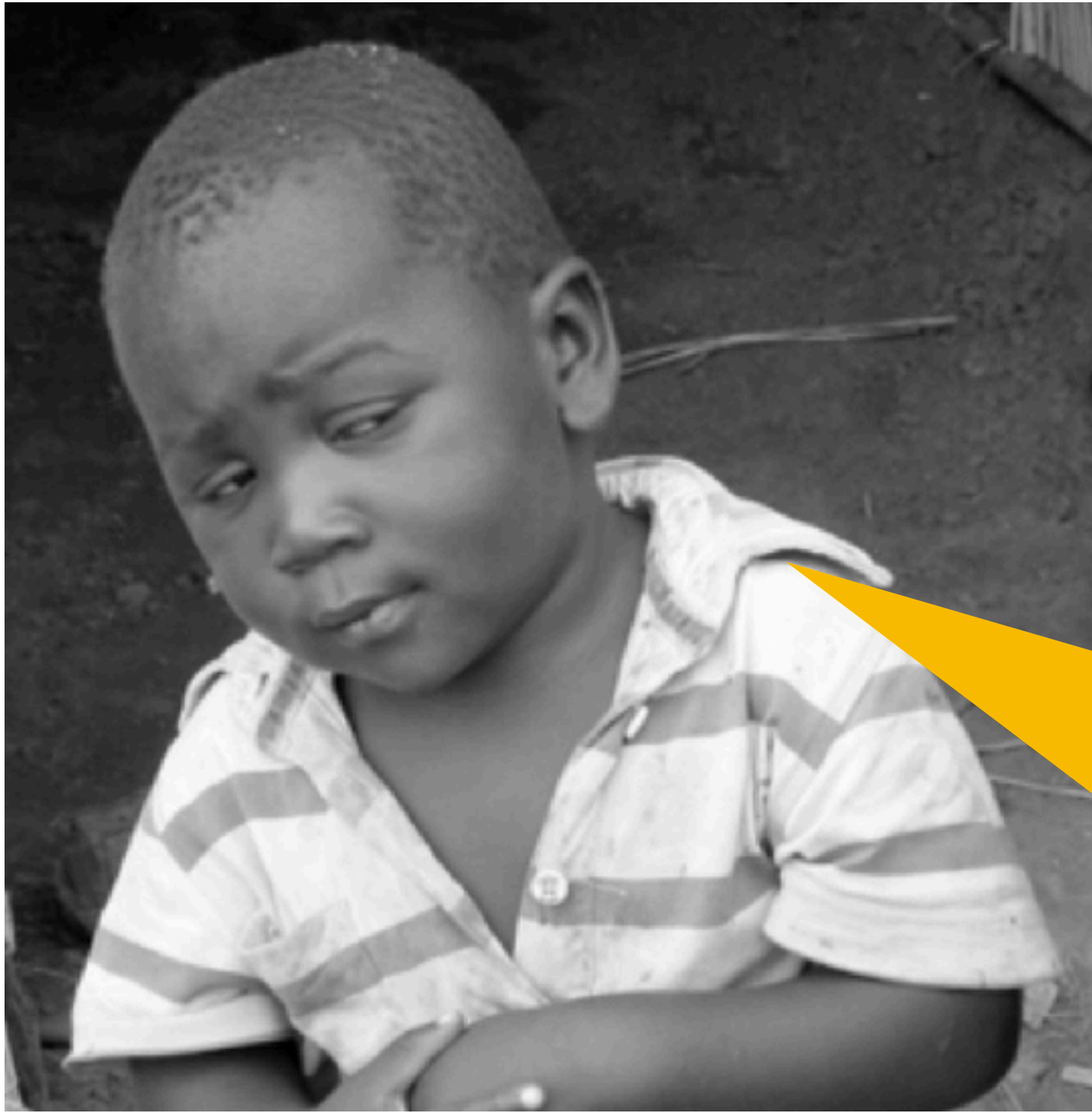
ජනාධිපතිවරණයට කලින් ප්‍රථිපල හරියටම කියූ හරුණයා (Pradeep Chanaka)

Raja Sapa · 735 views · 10 months ago

gotabaya_rajapaksha #sajith_premadasa #sri_lanka_politics#Election#result#prediction - Pradeep Chanakaඅපරාදේ වියදම් කරලා ජනදයක් පැවැත්තුවේ මෙයාගෙන් ඇහුවනම් ඉවරනේ.

**“I have predicted the
winner of the next
Presidential Election...”**





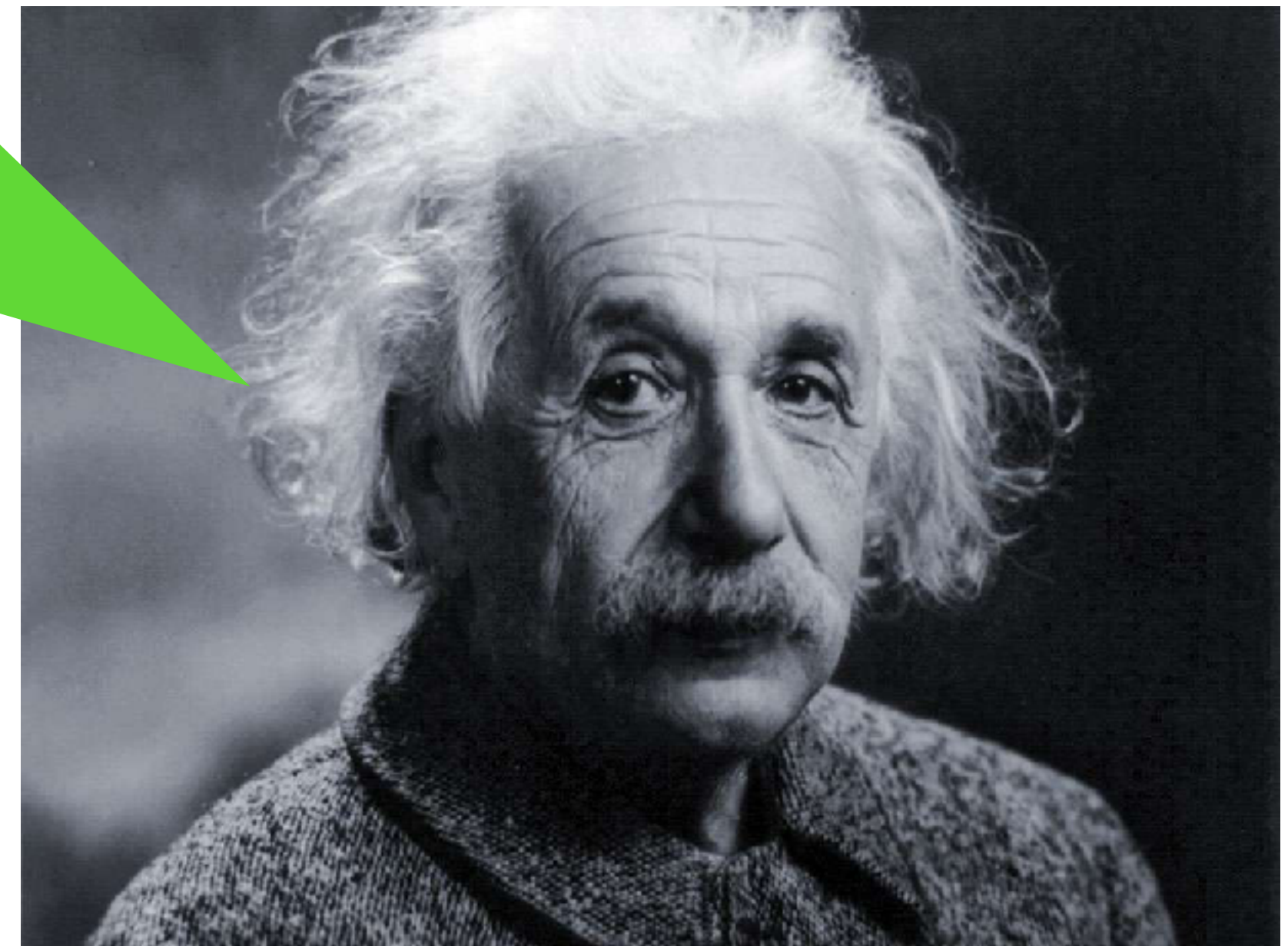
“Can you prove it?”

“I have predicted the winners of the two past elections...”

2015

2019

?

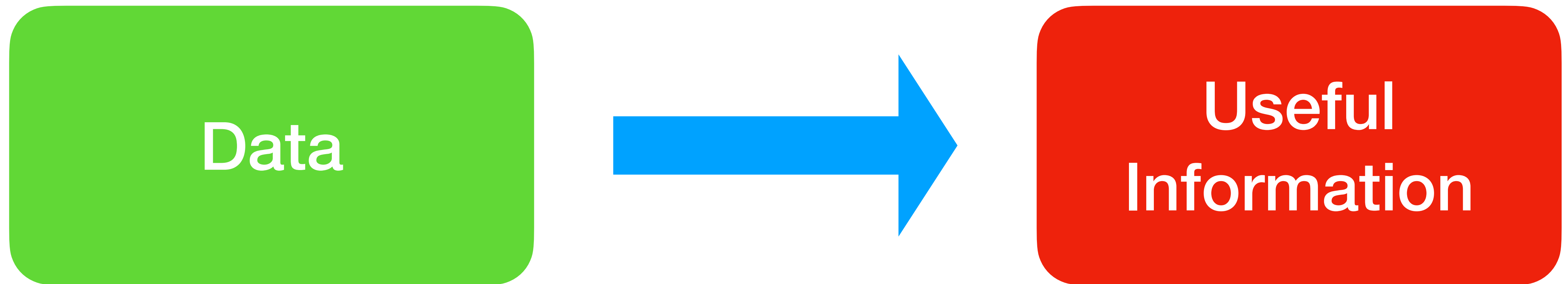


Data Science

“the use of **scientific methods** to
obtain **useful information** from
computer data,
especially large amounts of data”

<https://dictionary.cambridge.org>

Scientific Methods

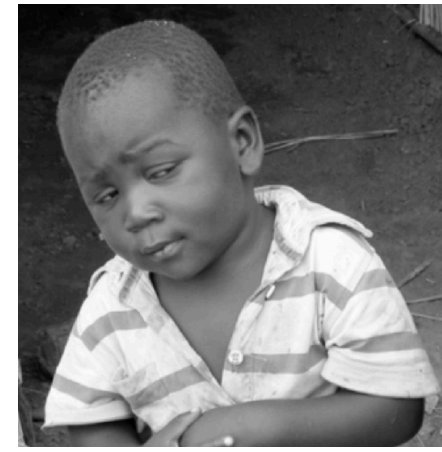
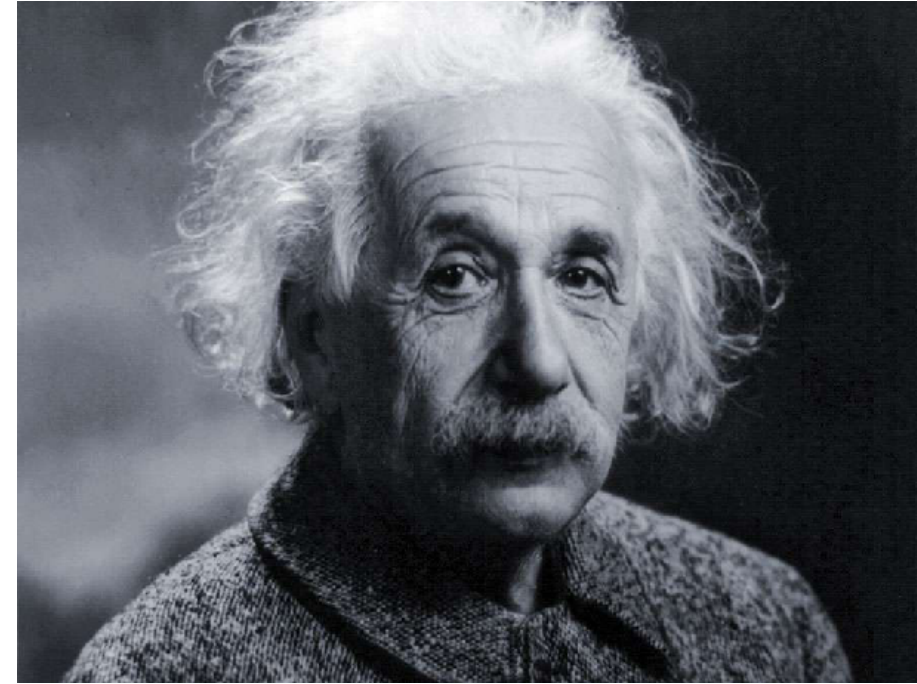


Statistics

Scientific Method



CASE 2



Basic Statistics

Scientific Method

“I predicted
2015 and 2019”

Data



“Can they correctly
predict the next
election?”

Useful Information

CASE 1

Statistical Confidence

“How **confident** we are about the **conclusions**
we derive from **data**”

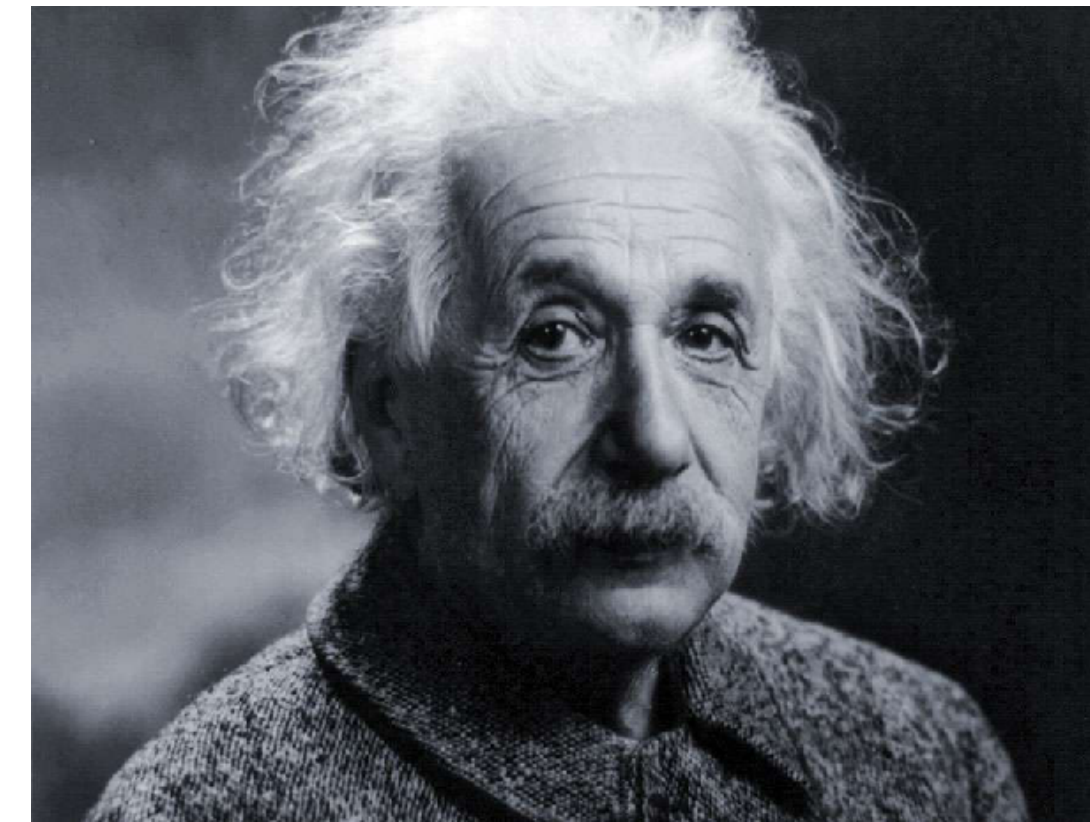
2015	2019	✓
2015	2019	✗
2015	2019	✗
2015	2019	✗





**Null
Hypothesis**

Fail



**Alternative
Hypothesis**

Pass

2015

2019

Test

New Vaccine
is **no better** than the
Placebo

Null
Hypothesis

Fail

New Vaccine
is **better** than the
Placebo

Alternative
Hypothesis

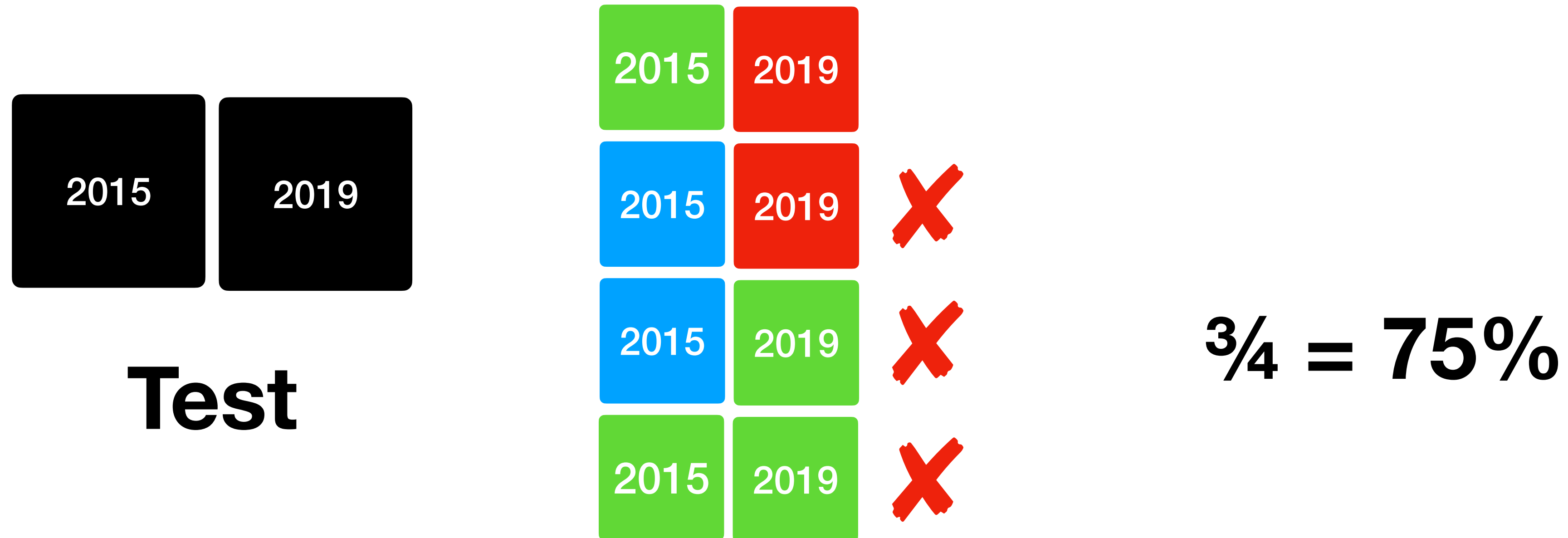
Pass

Vaccine Trials

Test

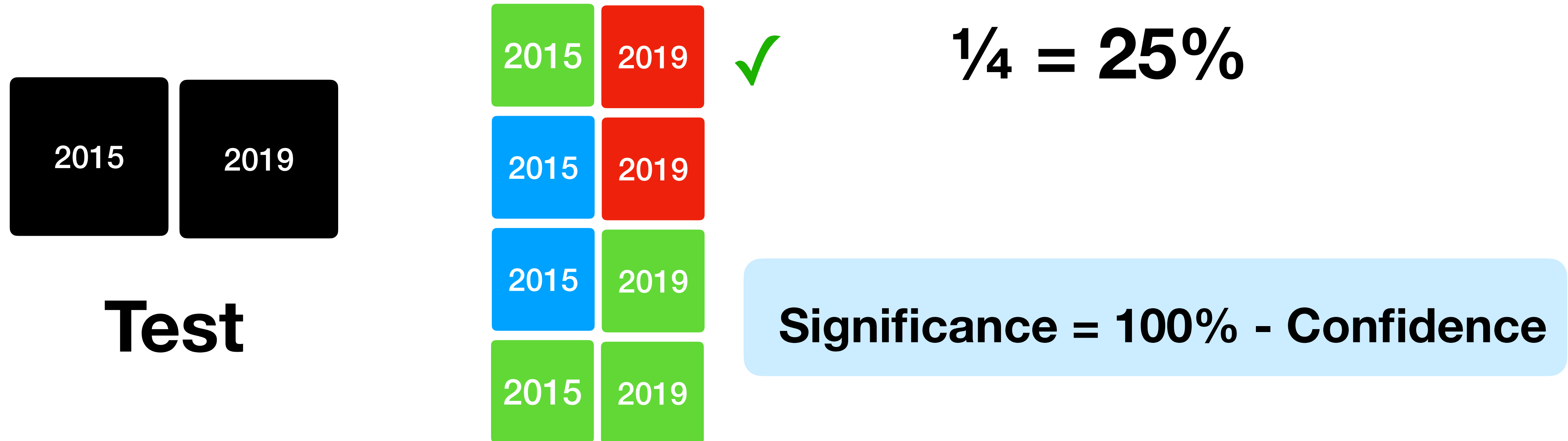
Statistical Confidence

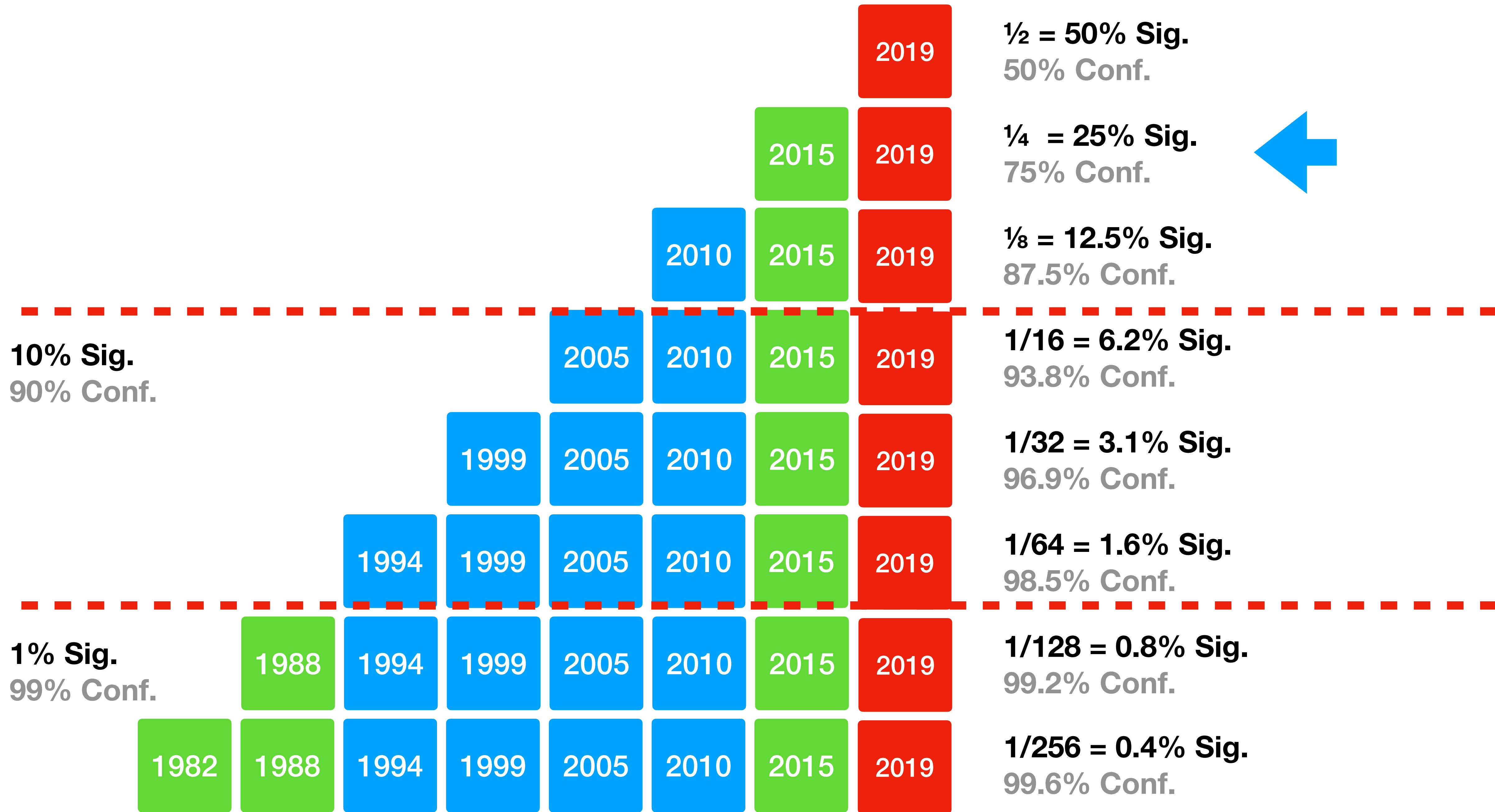
The Probability of **Failing The Test**, given that the **Null Hypothesis is actually true**



Statistical Significance

The Probability of **Passing The Test**, given that the **Null Hypothesis is actually true**

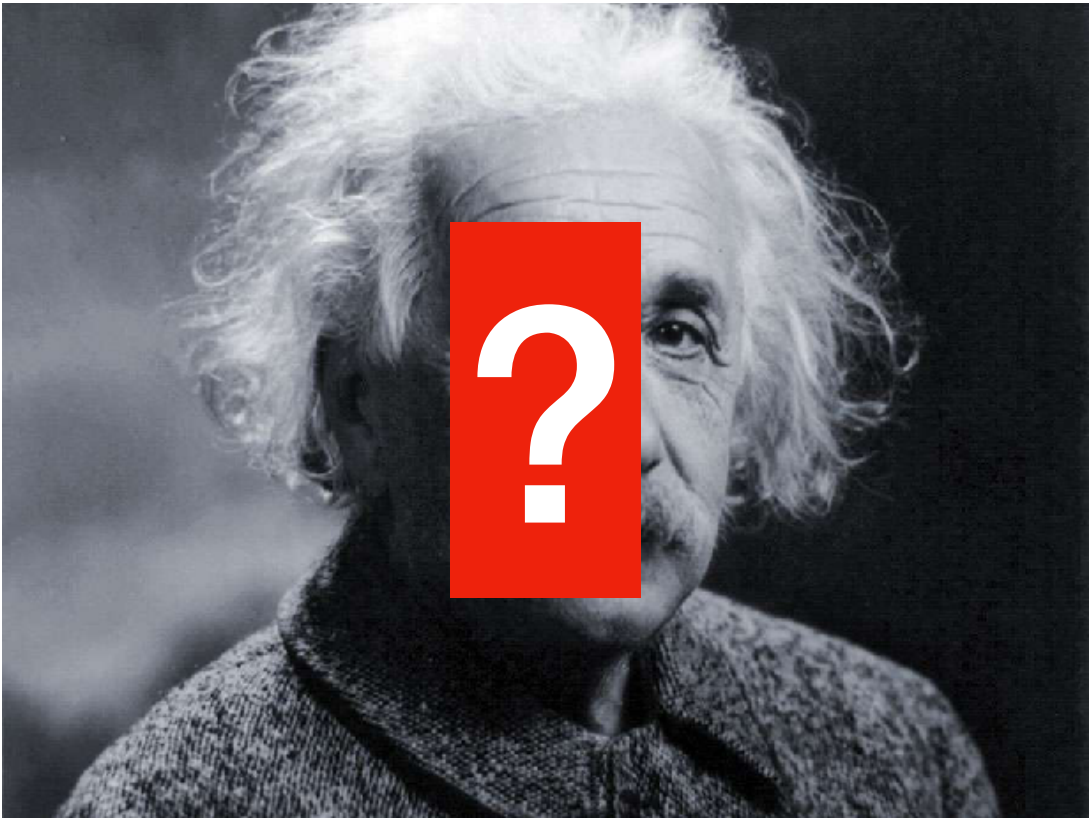




2015

2019

?



FACT-CHECK QUESTION 1

**What is the
Statistical Confidence?**

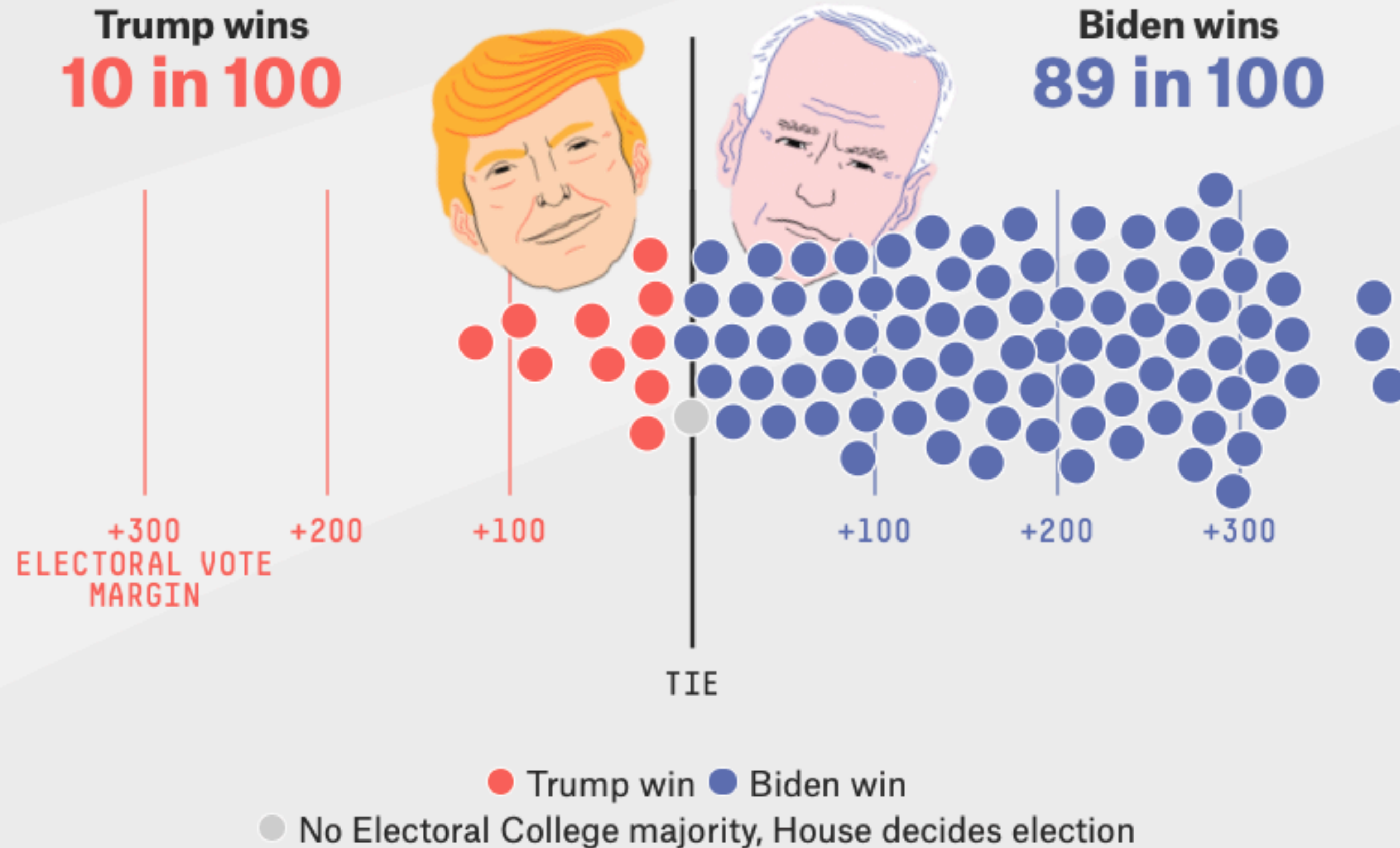


CASE 2

How to **Predict** Elections

Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.



<https://projects.fivethirtyeight.com/2020-election-forecast/>

Election Polling

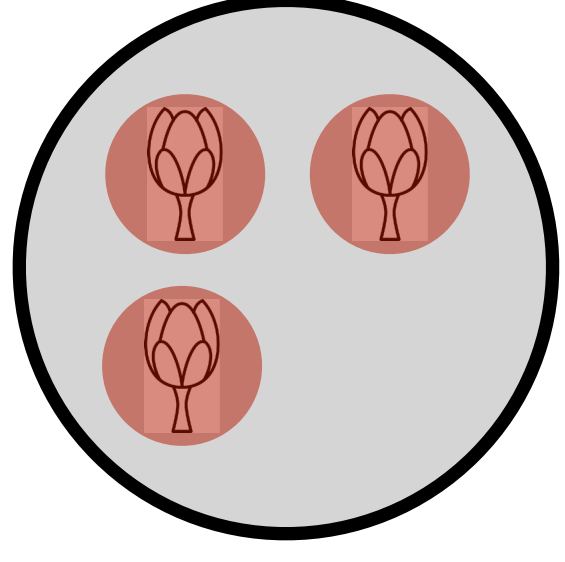
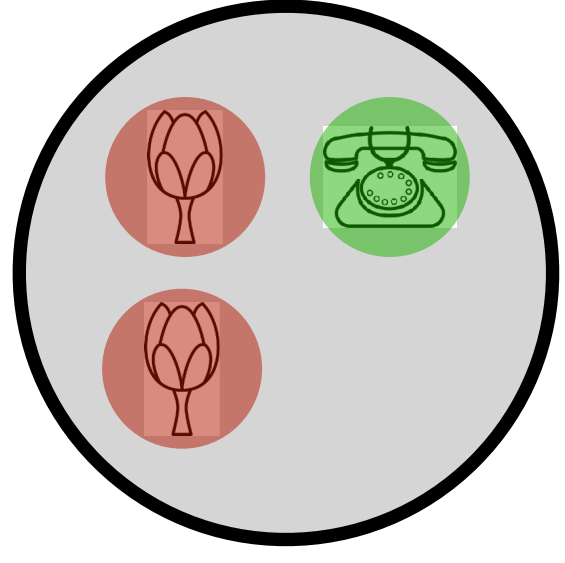
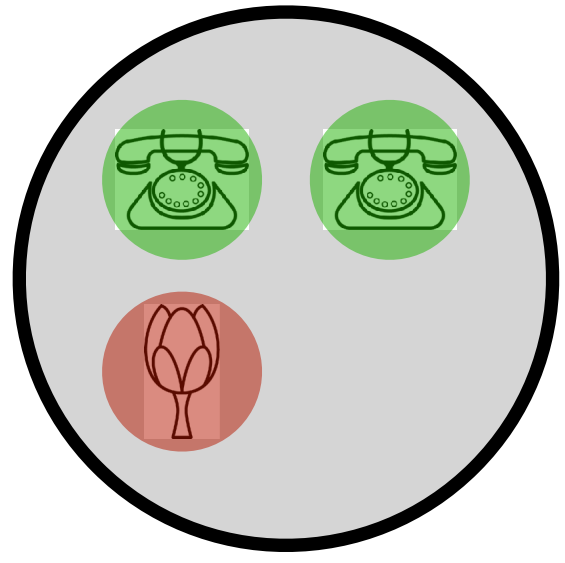
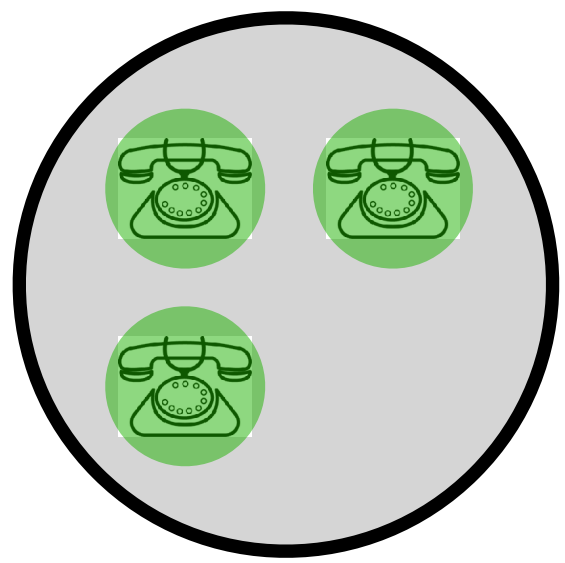
State-Level Presidential Head-to-Head Tracking

Respondents are asked, "If the November 2020 presidential election were being held today, for whom would you vote?"

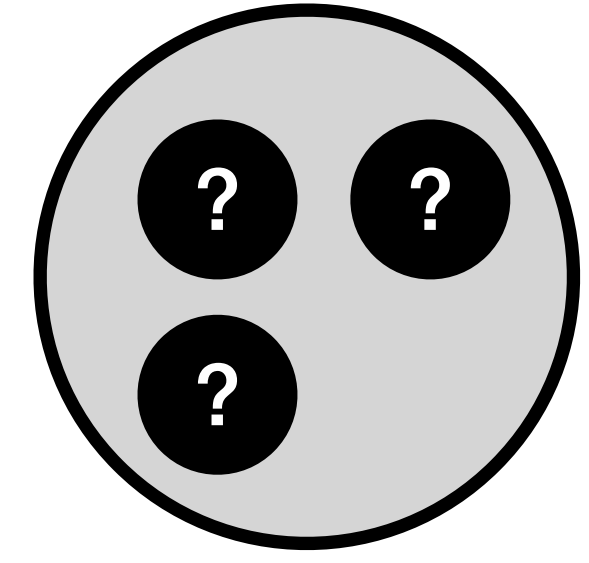
Each data point reflects surveys conducted between October 22-31, 2020 with likely voters.

	MARGIN	BIDEN %	TRUMP %
Arizona N: 1,059 MOE: 3%	Biden +2 Within MOE	48.4%	46.1%
Colorado N: 727 MOE: 4%	Biden +13	53.8%	40.9%
Florida N: 4,451 MOE: 2%	Biden +6	51.5%	45.4%
Georgia N: 1,743 MOE: 2%	Biden +2 Within MOE	48.8%	46.4%
Indiana N: 1,147 MOE: 3%	Trump +11	42.0%	53.4%
Michigan N: 1,736 MOE: 2%	Biden +7	51.8%	44.5%

<https://morningconsult.com>



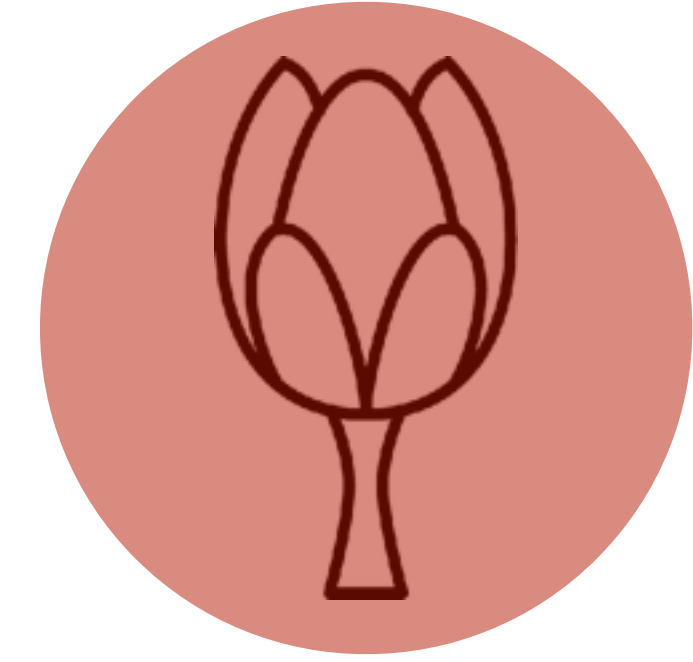
4 Groups



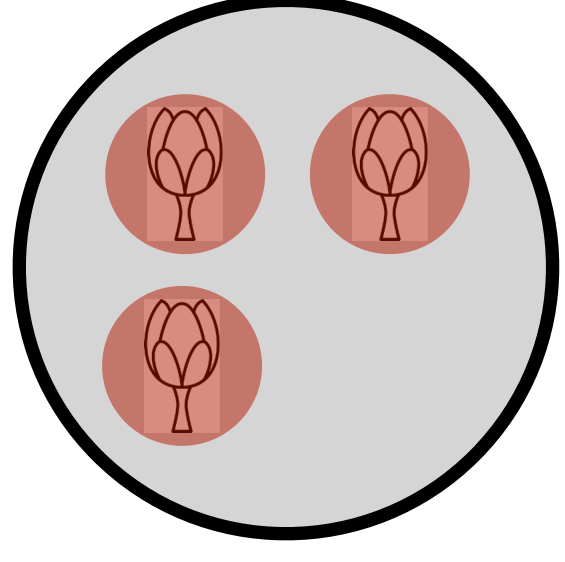
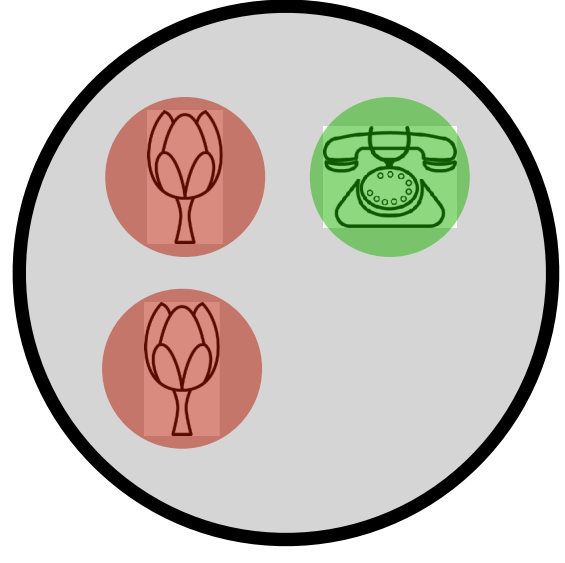
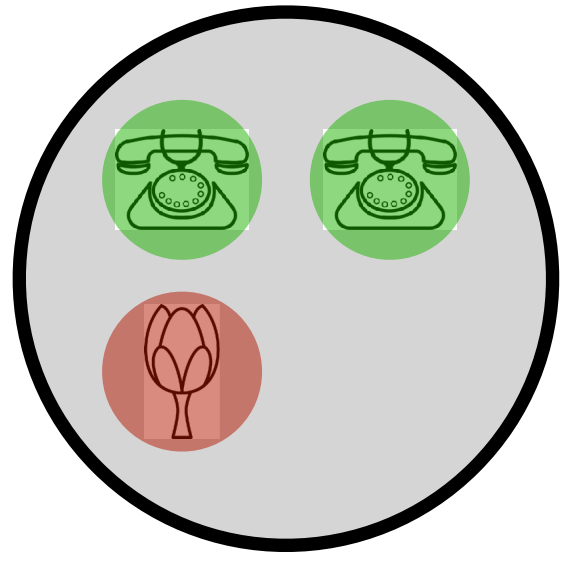
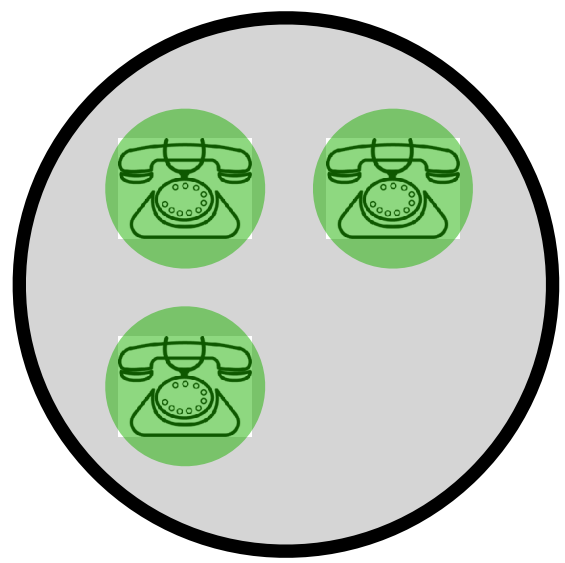
Random Pick



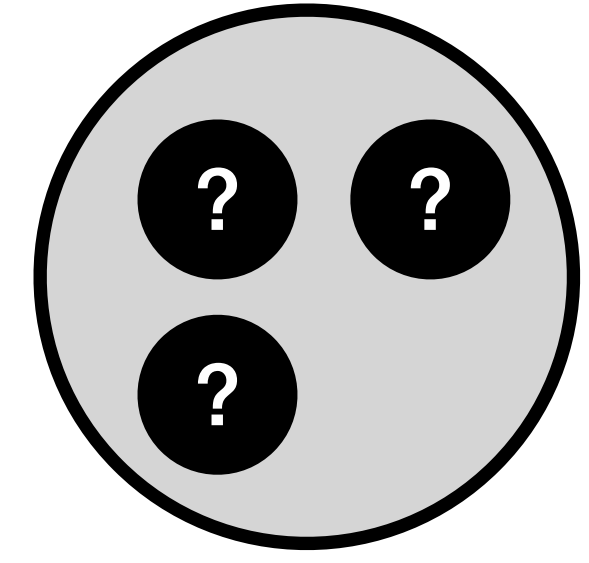
SJB



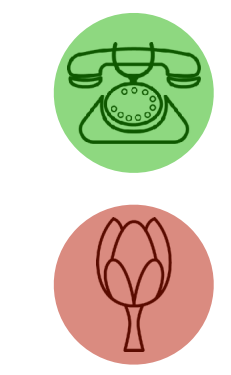
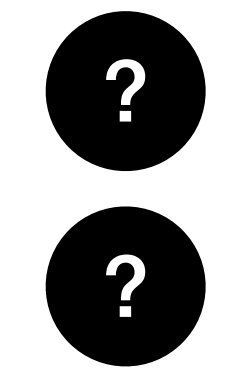
SLPP



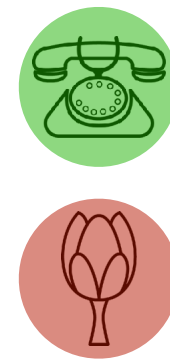
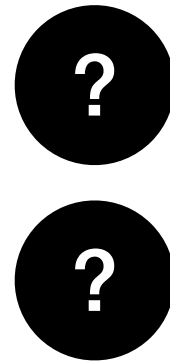
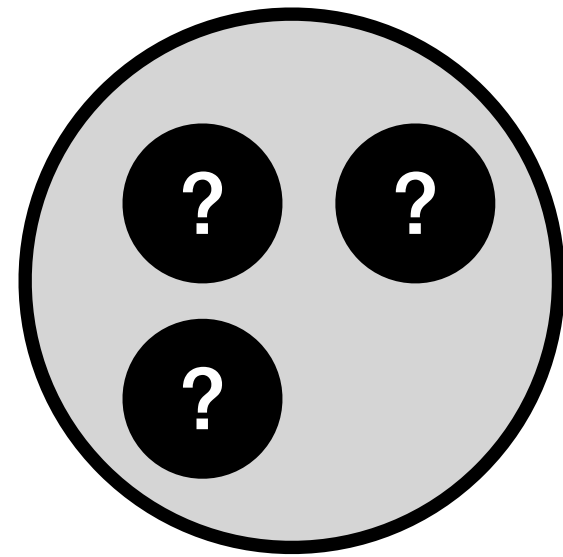
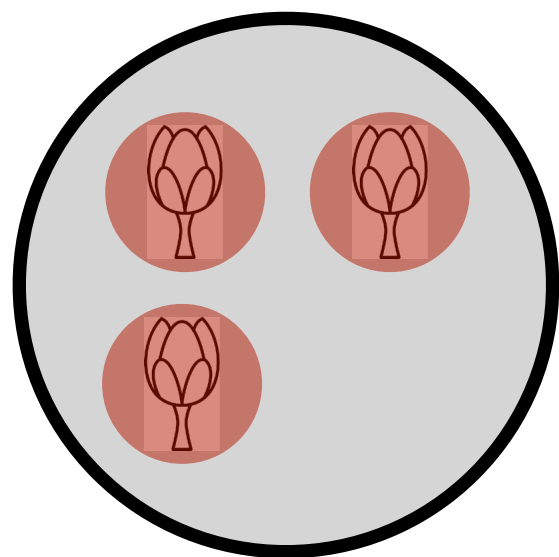
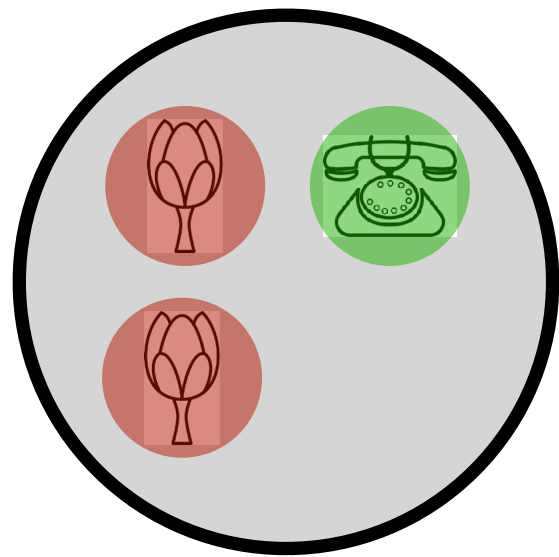
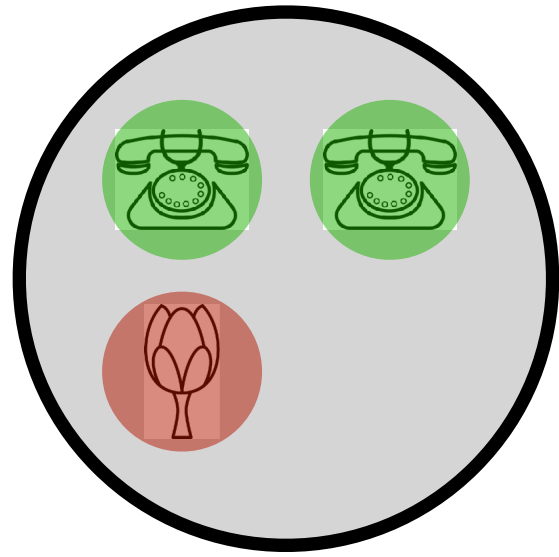
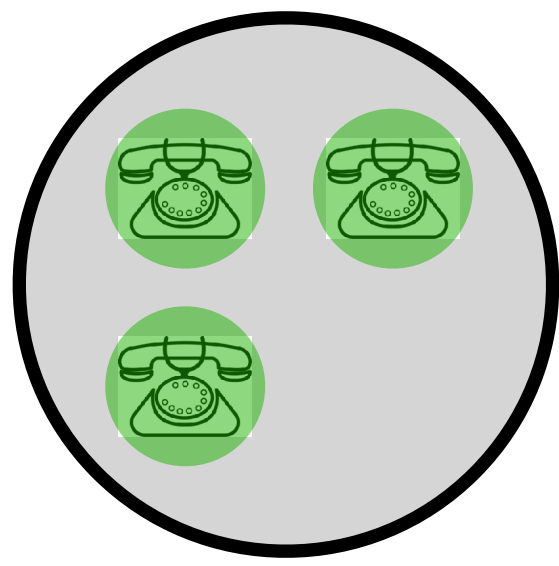
4 Groups



Random Pick



2 People Randomly picked from the group

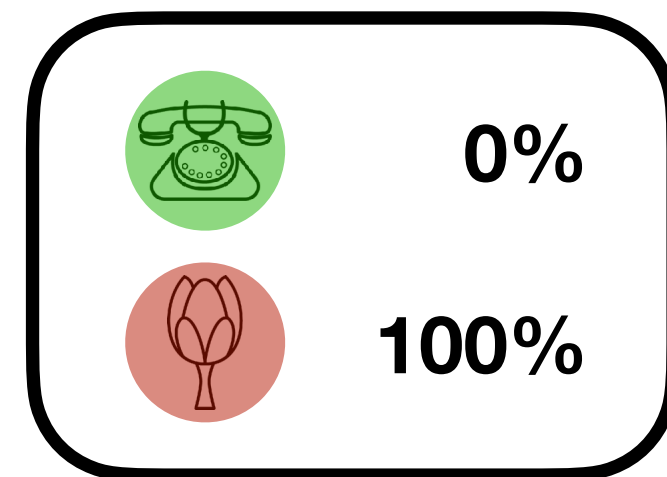
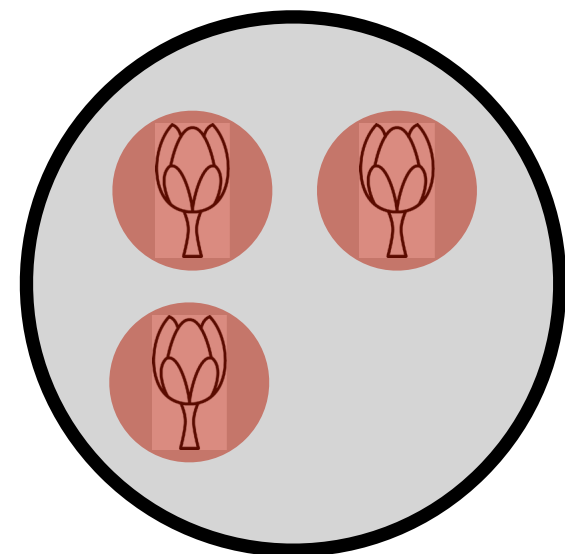
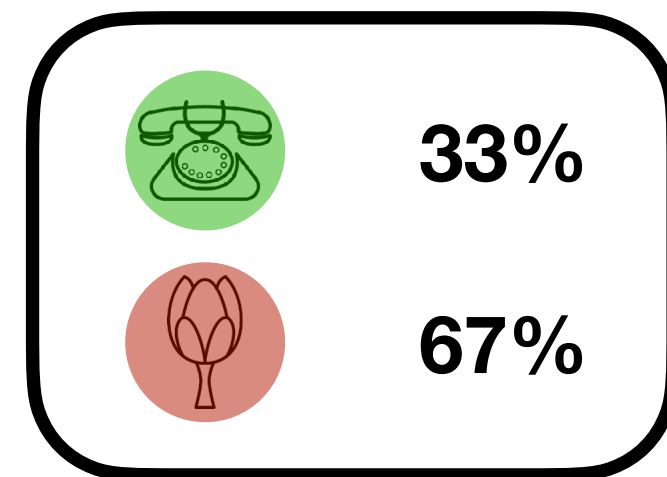
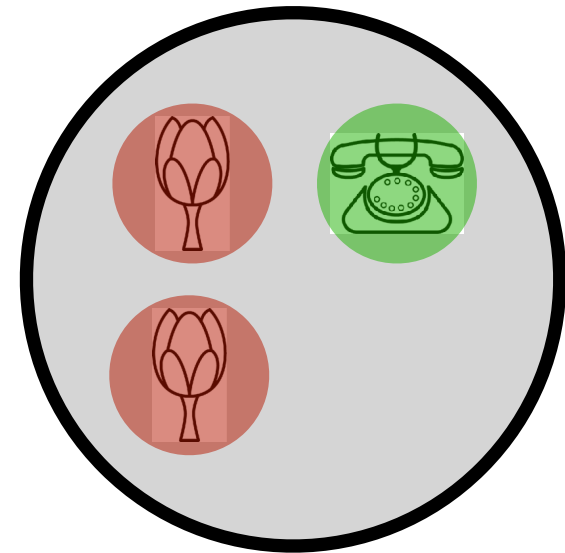
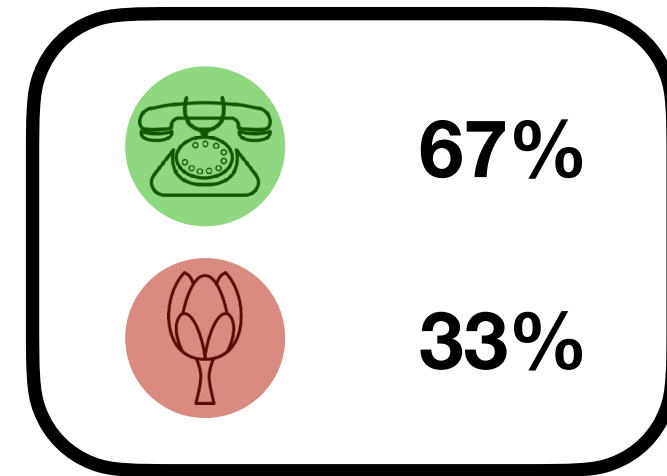
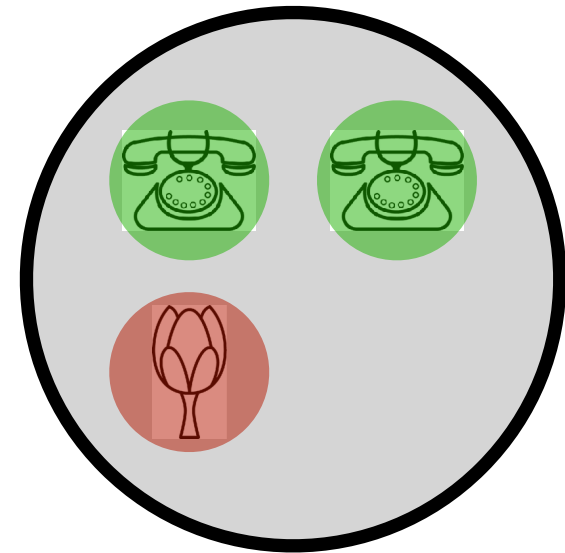
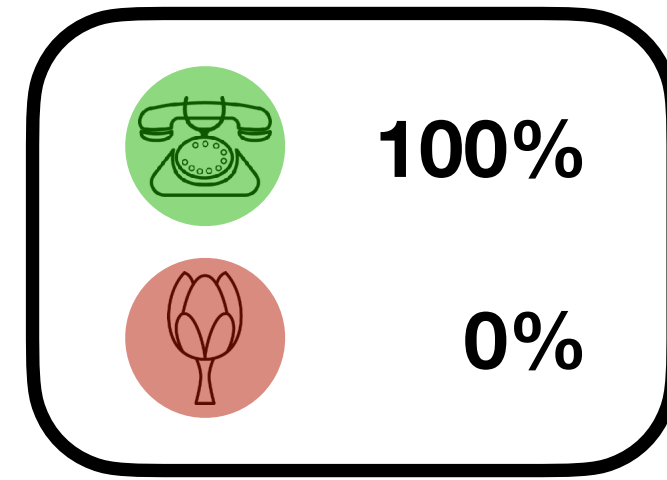
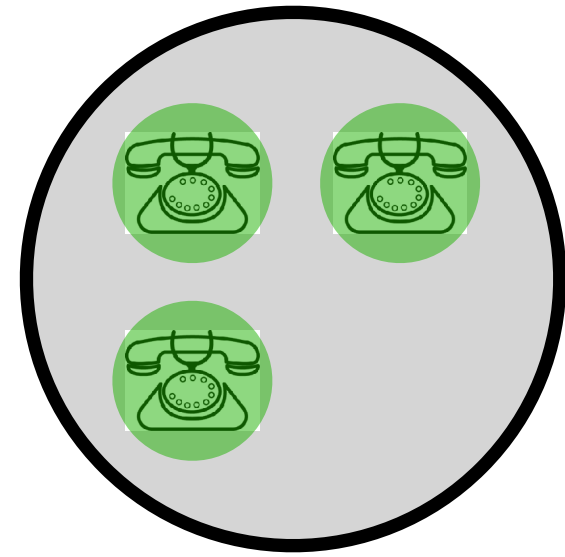


Potential
Voter Populations

Actual
Voter Population

Sample

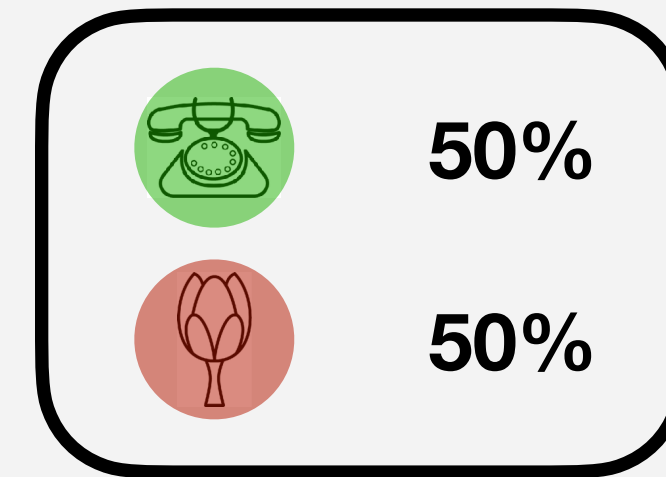
Population



50%

50%

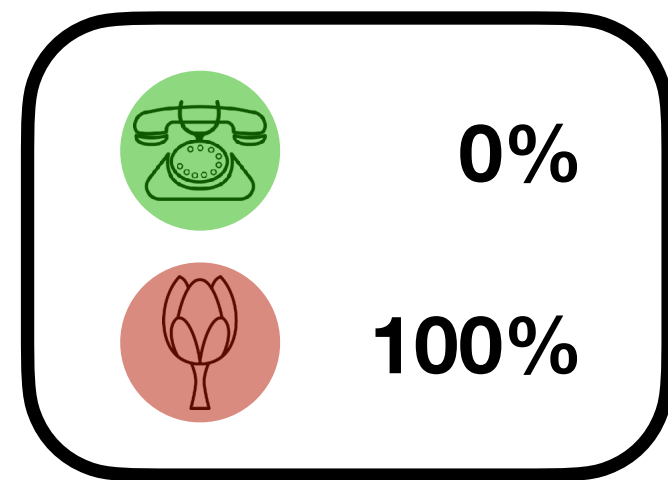
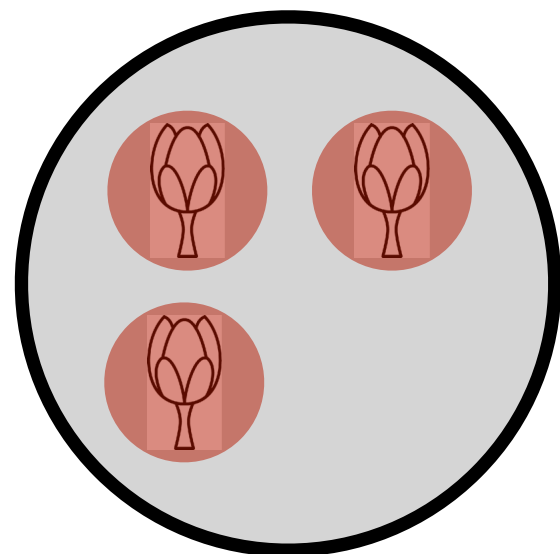
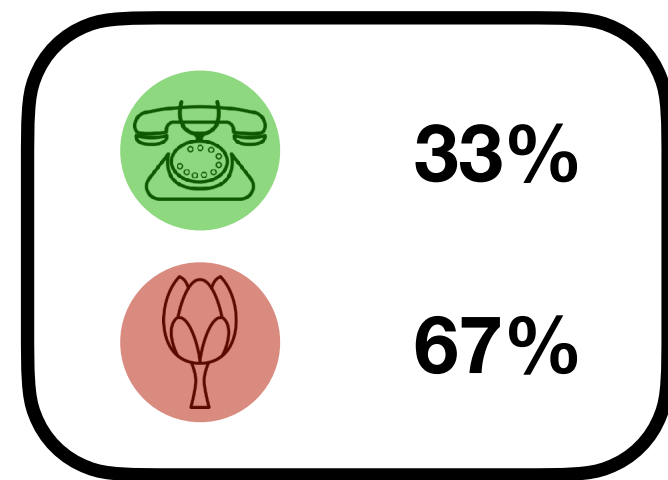
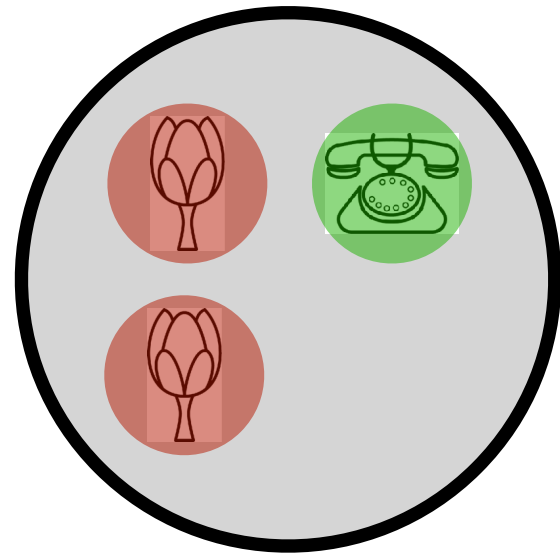
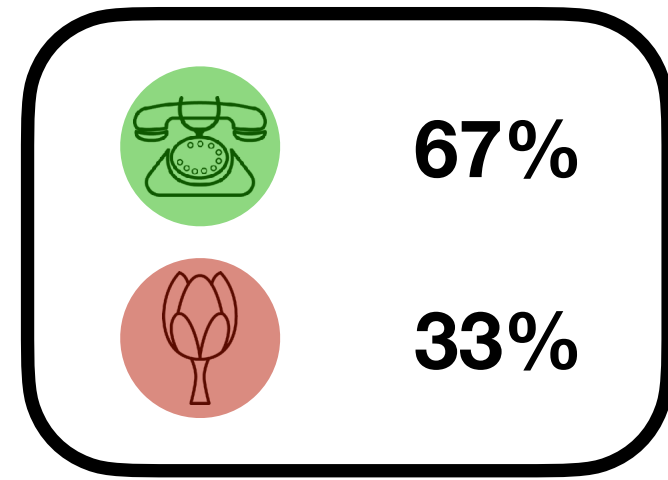
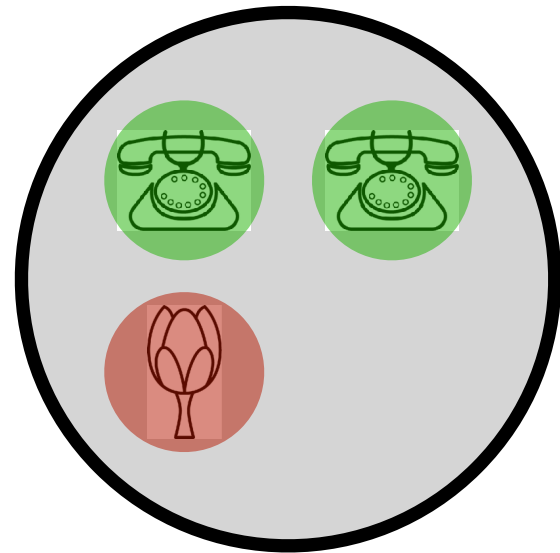
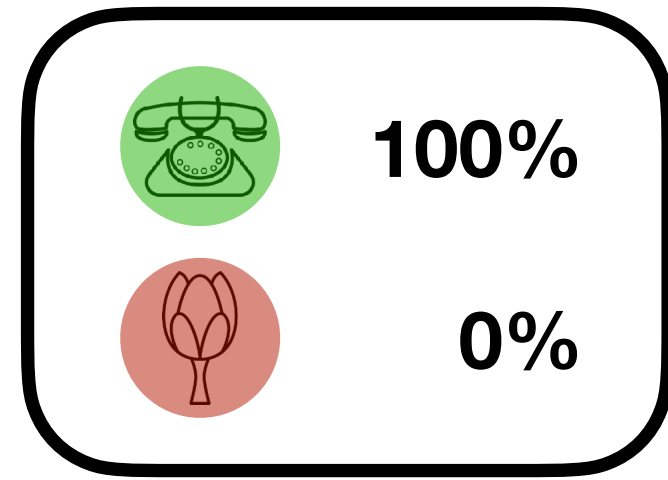
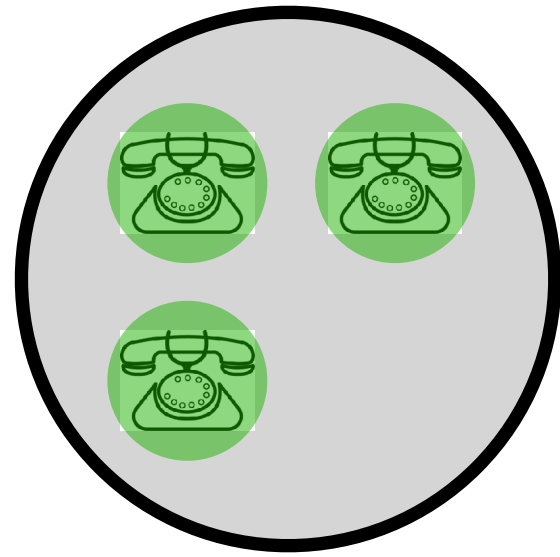
Sample



Sampling Error

“Characteristics of the **Sample** are not equal to characteristics of the **Population**”

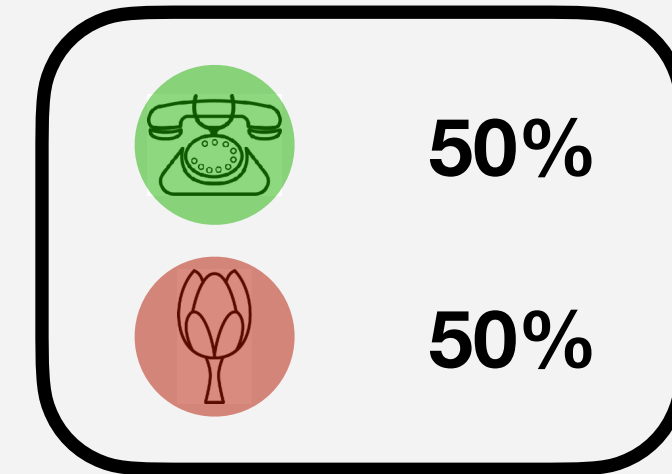
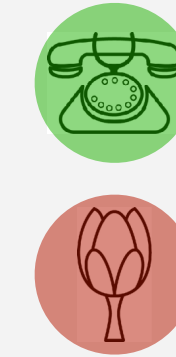
Population



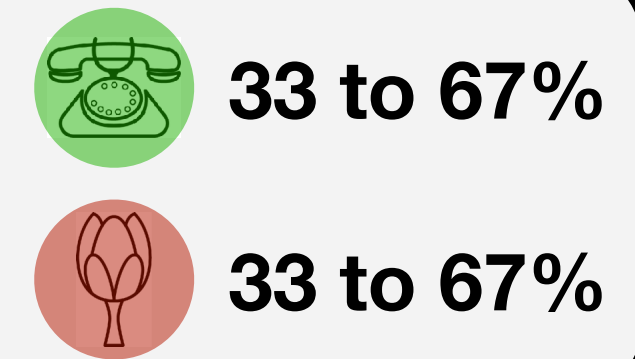
50%

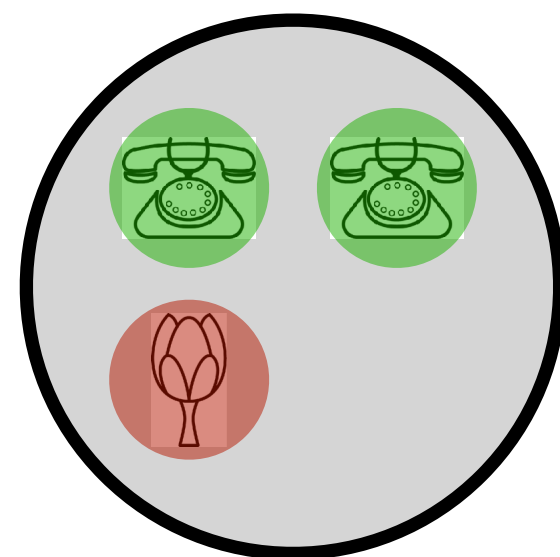
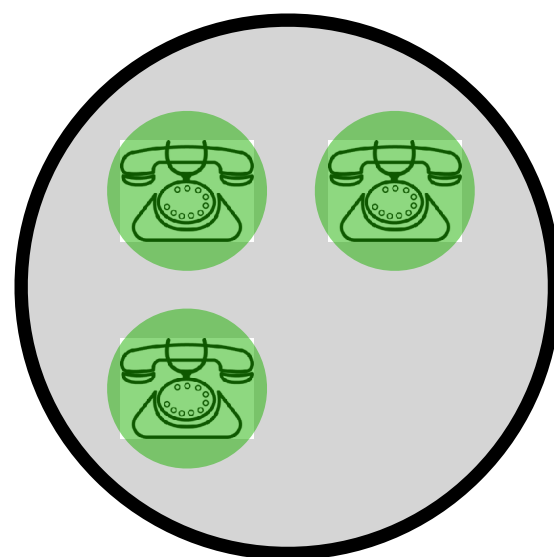
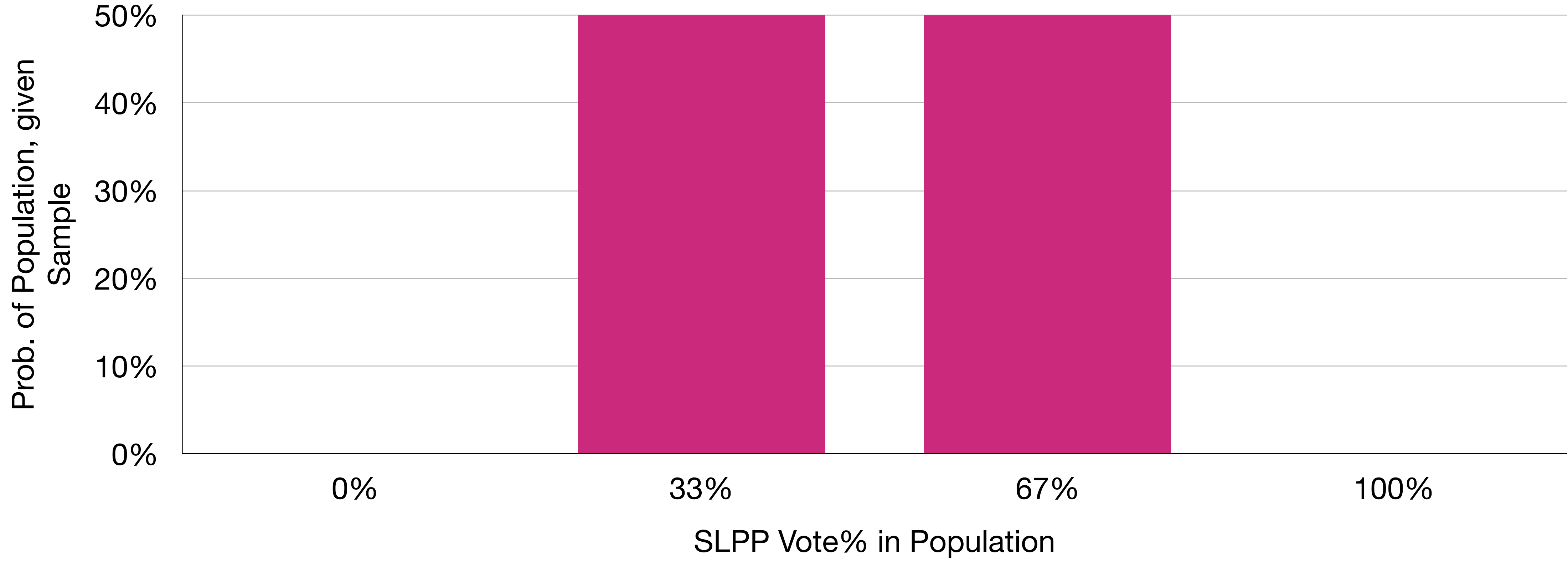
50%

Sample

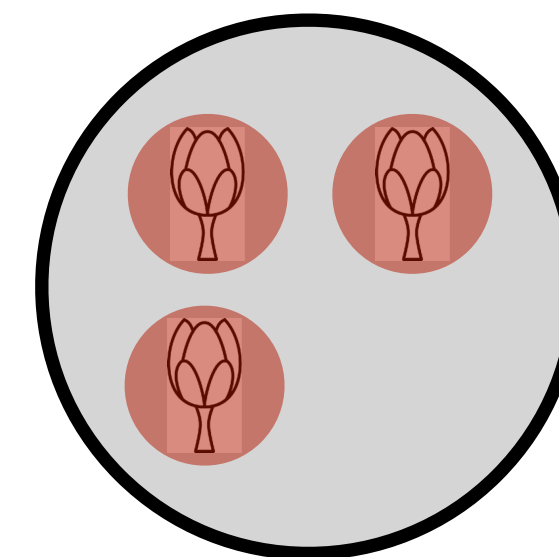
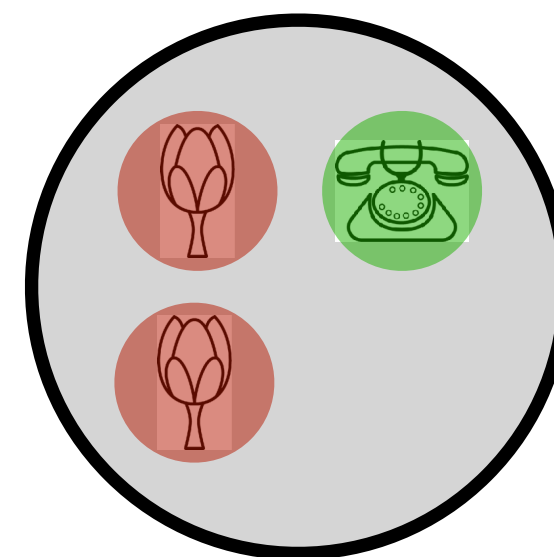


Conclusions





33

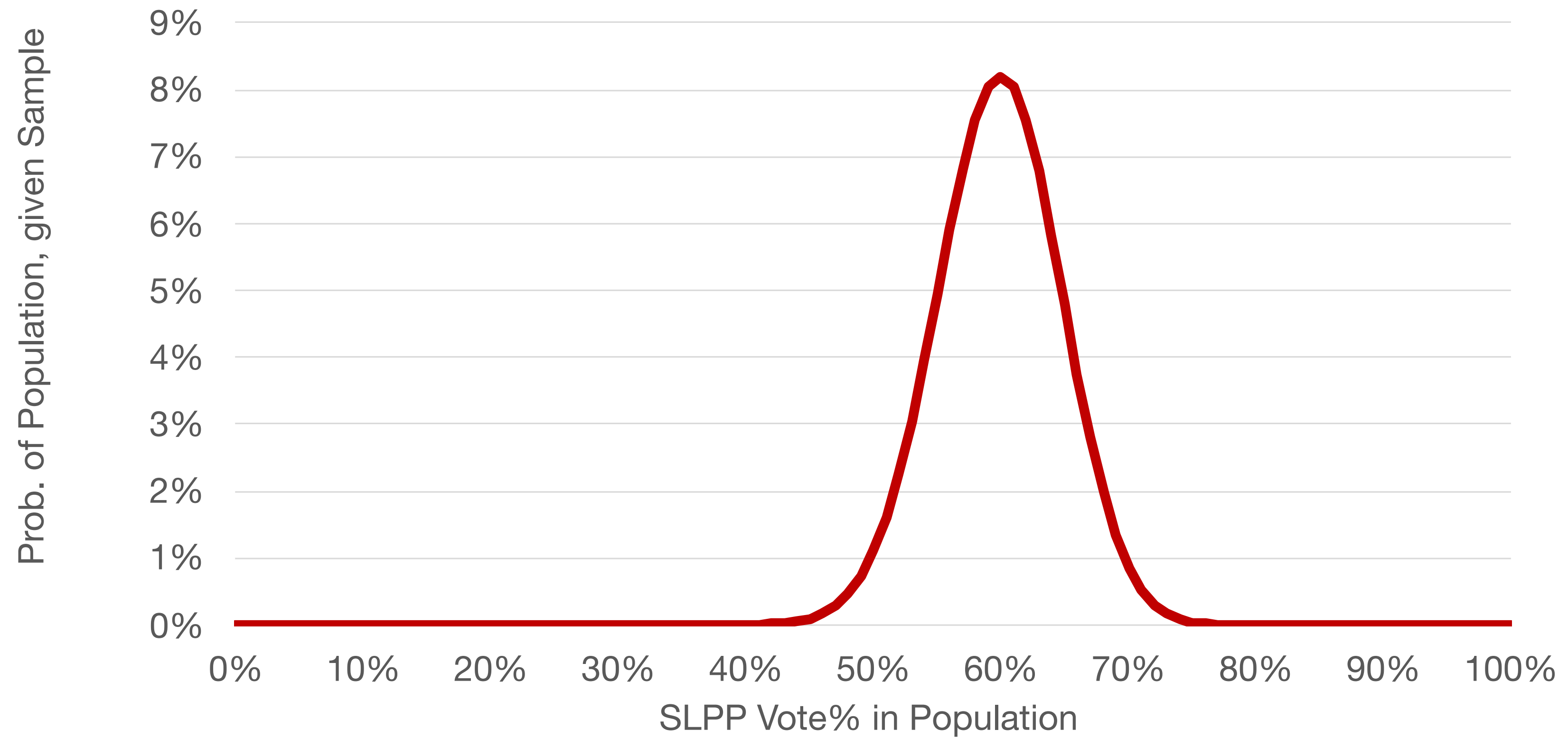
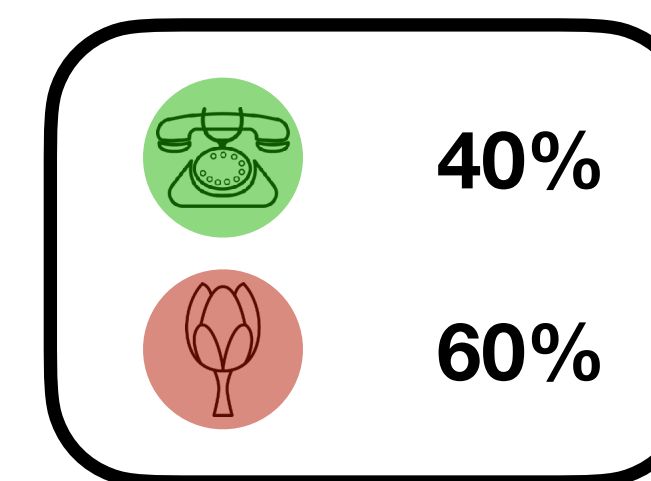


Population

(Colombo ED)

Sample

(n = 100)

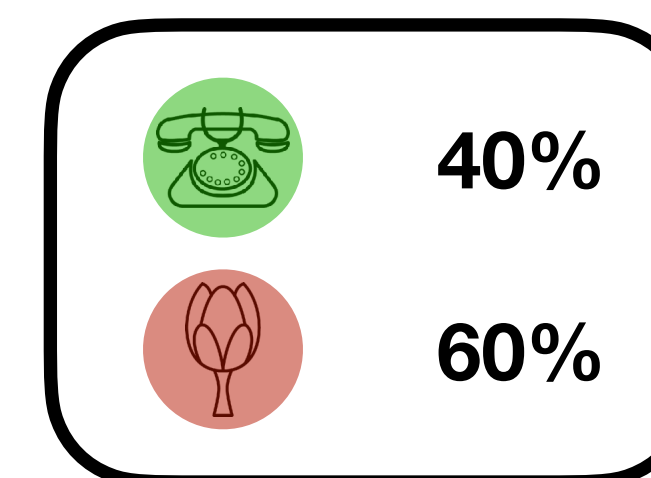


Population

(Colombo ED)

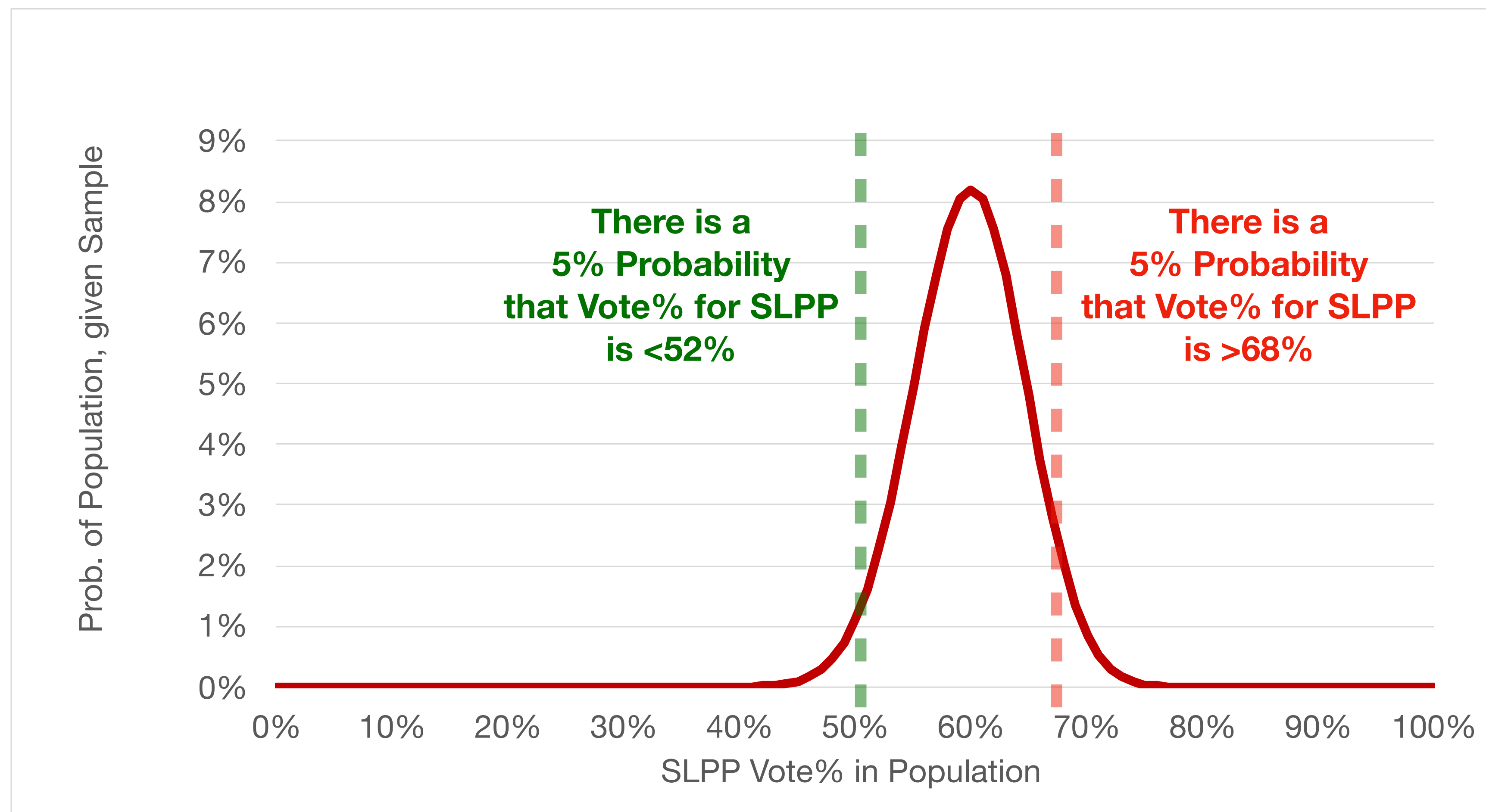
Sample

(n = 100)

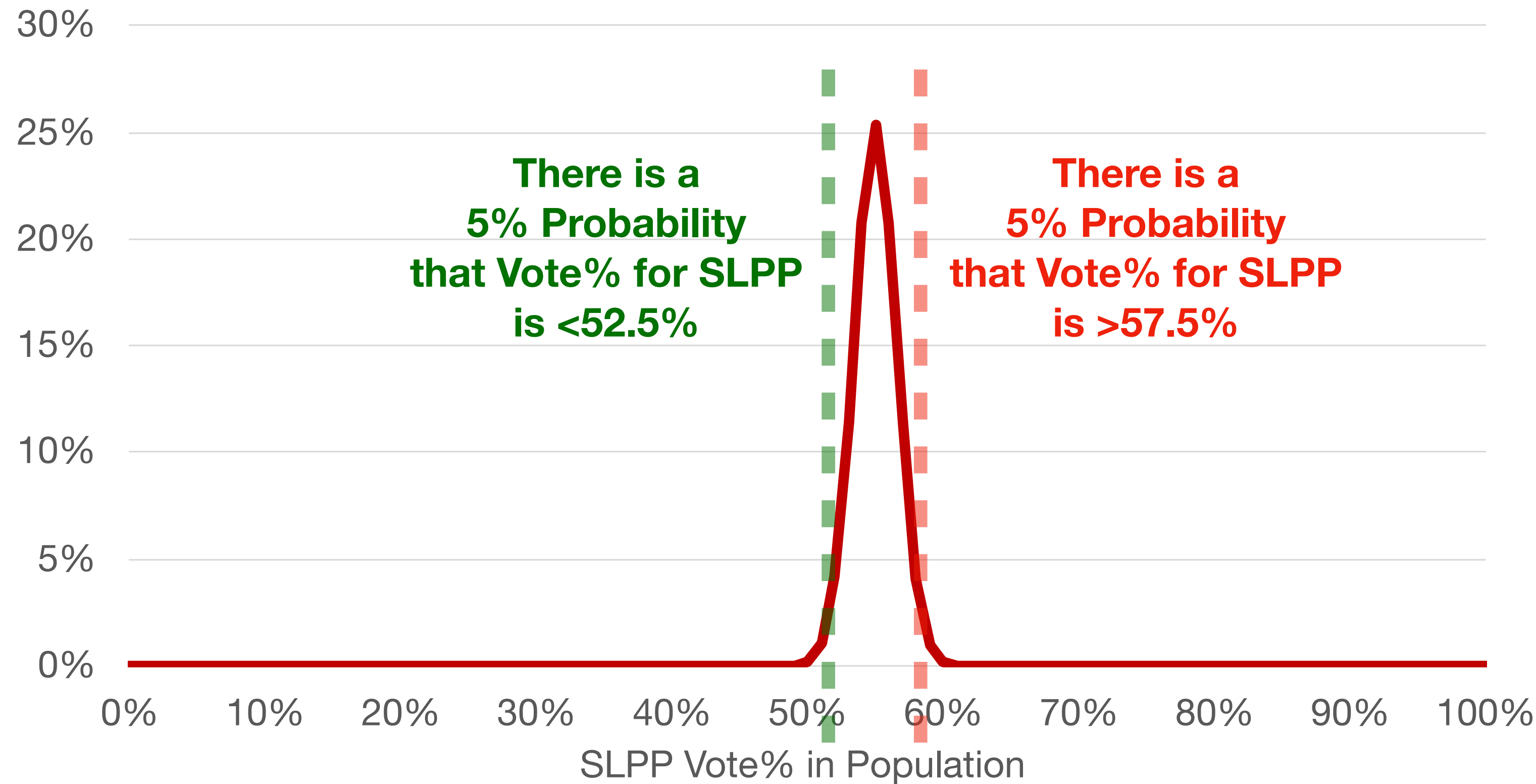


There is a
90% Probability
that **Vote% for SLPP**
is between **52% and 68%**

52% to 68%
60% ± 8%

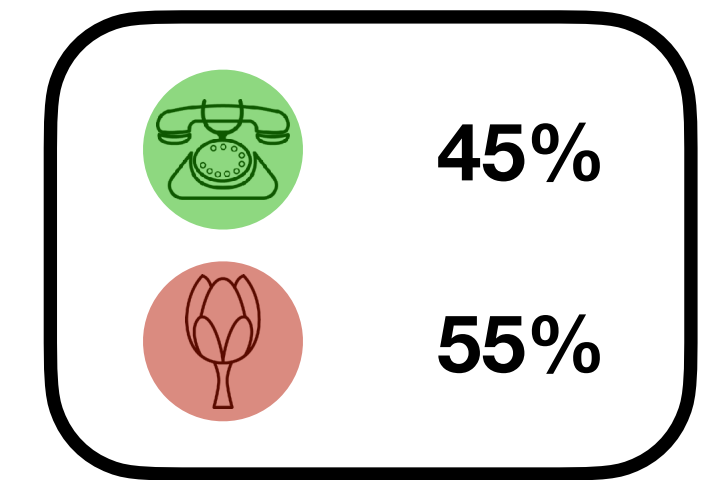


Prob. of Population, given Sample



Sample

(n = 1,000)



There is a
90% Probability
that **Vote% for SLPP**
is between **52.5% and 57.5%**

52.5% to 57.5%
55% ± 2.5%

FACT-CHECK QUESTION 2

**What are the
90% Confidence
Error Bounds?**

State-Level Presidential Head-to-Head Tracking

Respondents are asked, "If the November 2020 presidential election were being held today, for whom would you vote?"

Each data point reflects surveys conducted between October 22-31, 2020 with likely voters.

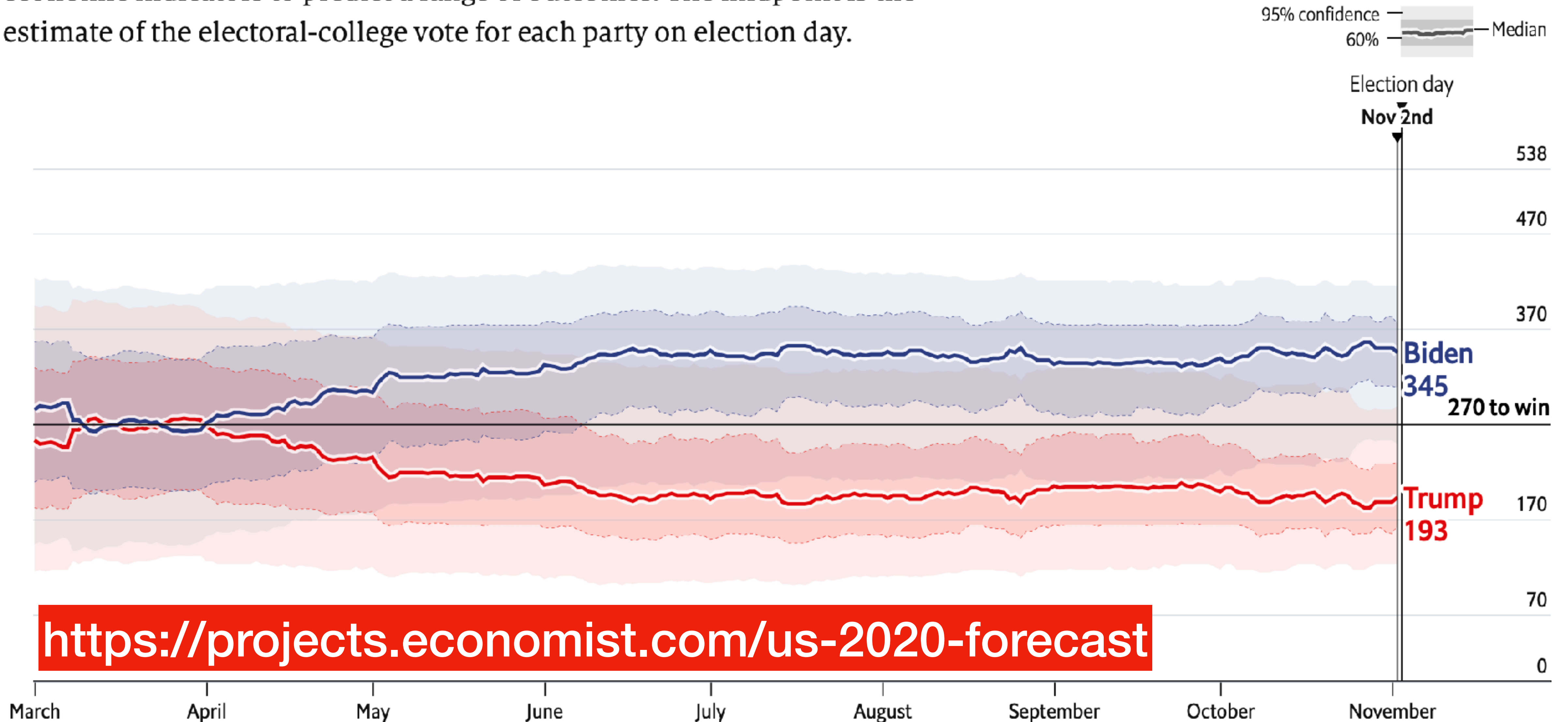
	MARGIN	BIDEN %	TRUMP %
Arizona N: 1,059 MOE: 3%	Biden +2 Within MOE	48.4%	46.1%
Colorado N: 727 MOE: 4%	Biden +13	53.8%	40.9%
Florida N: 4,451 MOE: 2%	Biden +6	51.5%	45.4%
Georgia N: 1,743 MOE: 2%	Biden +2 Within MOE	48.8%	46.4%
Indiana N: 1,147 MOE: 3%	Trump +11	42.0%	53.4%
Michigan N: 1,736 MOE: 2%	Biden +7	51.8%	44.5%

**Biden could be
Behind by 1% point
to
Ahead by 5% points**

<https://morningconsult.com>

Estimated electoral college votes

Our model is updated every day and combines state and national polls with economic indicators to predict a range of outcomes. The midpoint is the estimate of the electoral-college vote for each party on election day.



<https://projects.economist.com/us-2020-forecast>

Colombo Electoral District

2020 Parliamentary Election



Predicted

9 to 12 seats

6 to 8 seats

0 to 2 seats

0 to 2 seats

90% Confidence

Sample n = 300

Colombo Electoral District

2020 Parliamentary Election



9 to 12 seats

6 seats



6 to 8 seats

12 seats



0 to 2 seats

1 seats

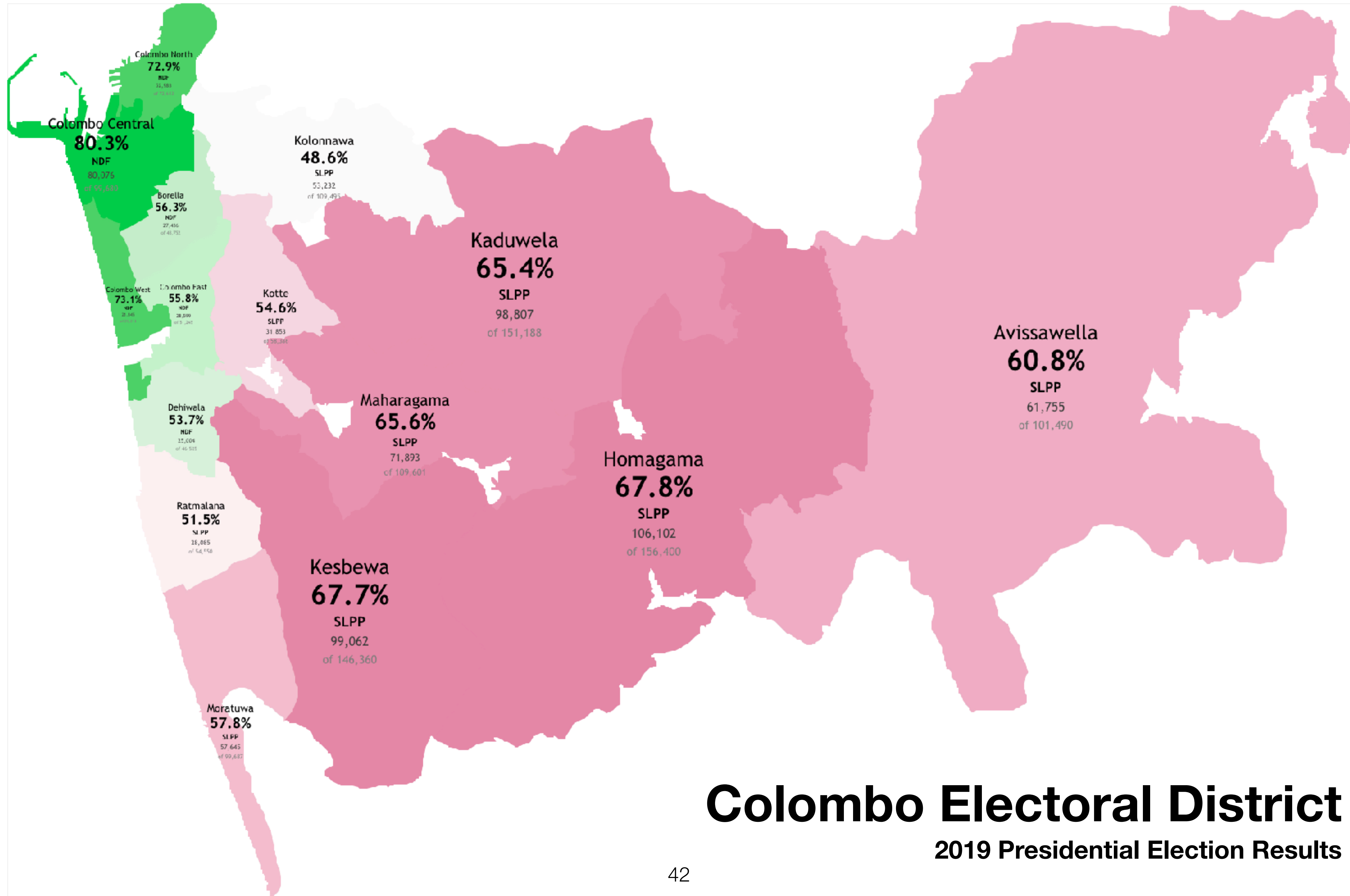


0 to 2 seats

0 seats

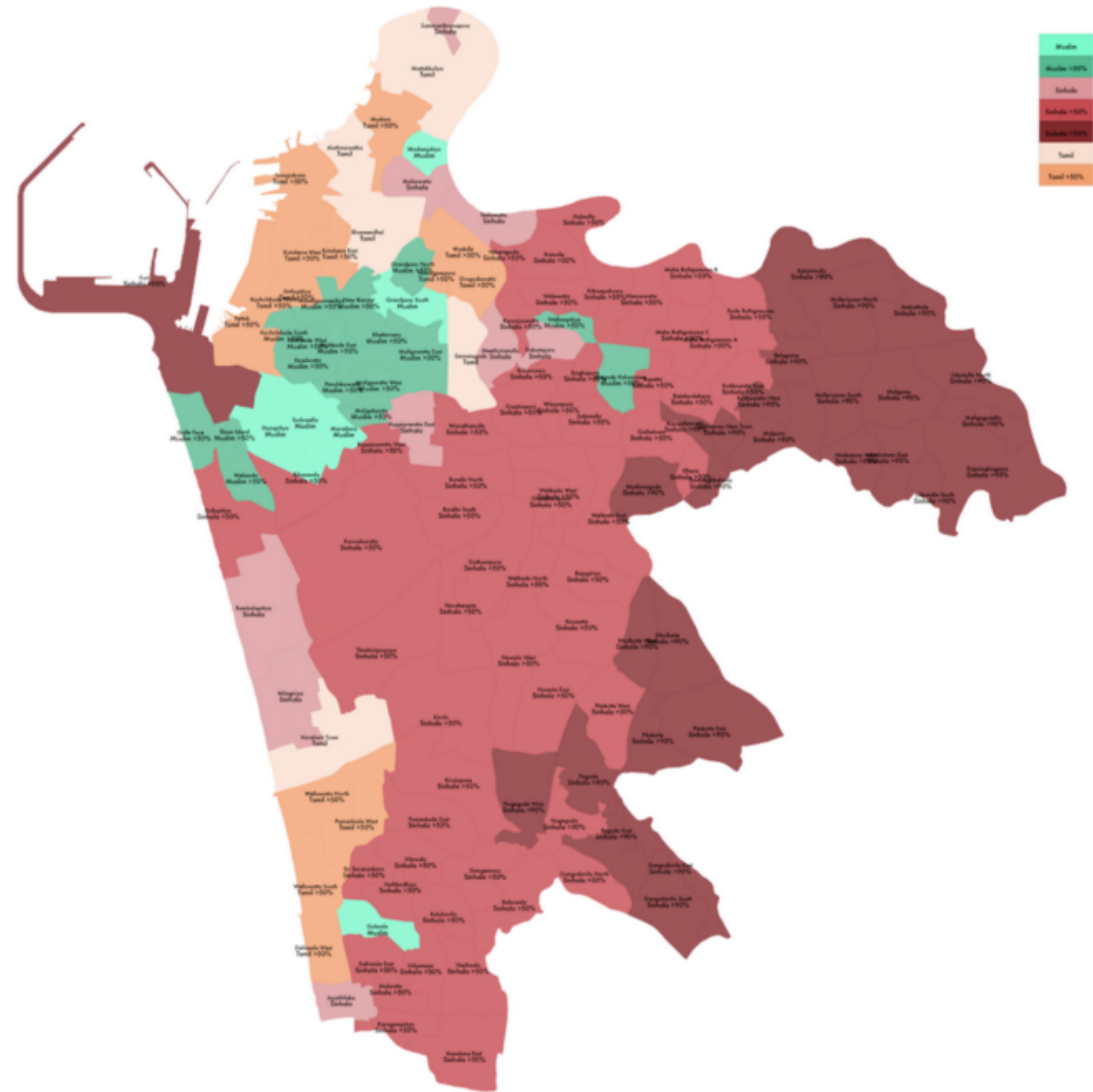


**90% Confidence
Sample n = 300**



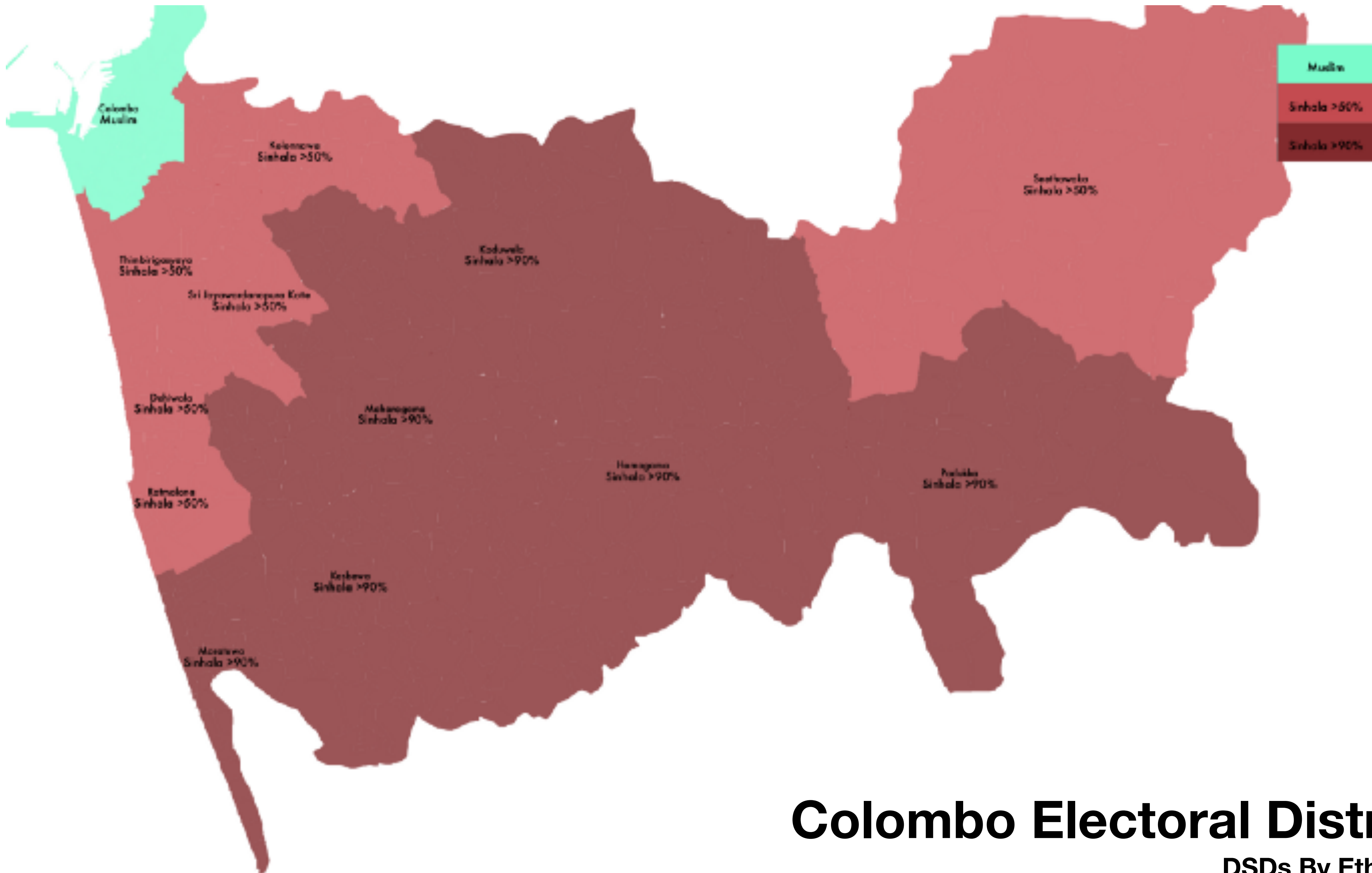
Colombo Electoral District

2019 Presidential Election Results



Colombo MC, Kotte, Dehiwala, Kolonnawa

DSDs By Ethnicity



Colombo Electoral District
DSDs By Ethnicity

FACT-CHECK QUESTION 3

Did you check if the
sample was **random**?



Facebook users in Sri Lanka

January 2019



6 168 000



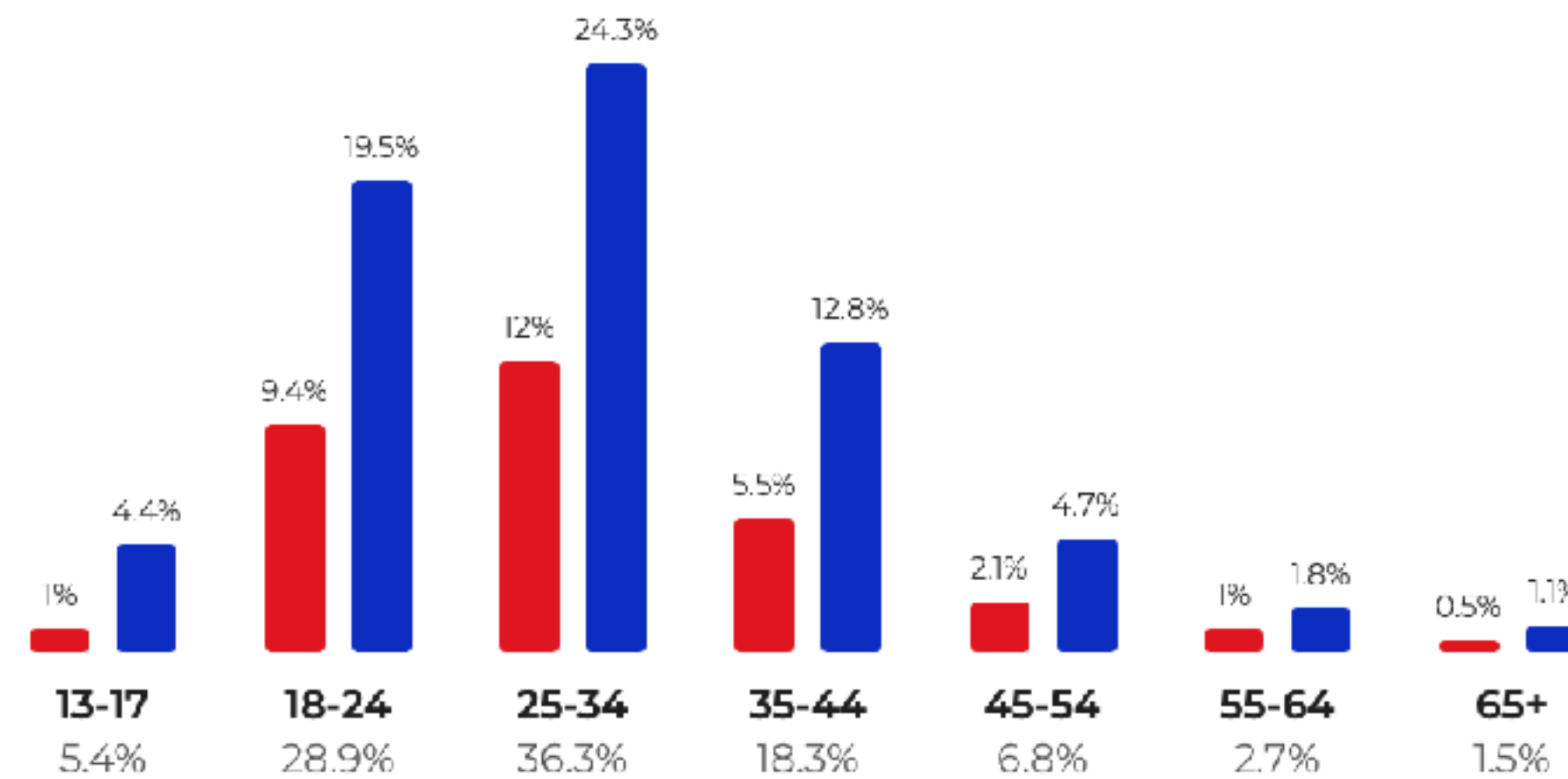
31.5%

women



68.5%

men



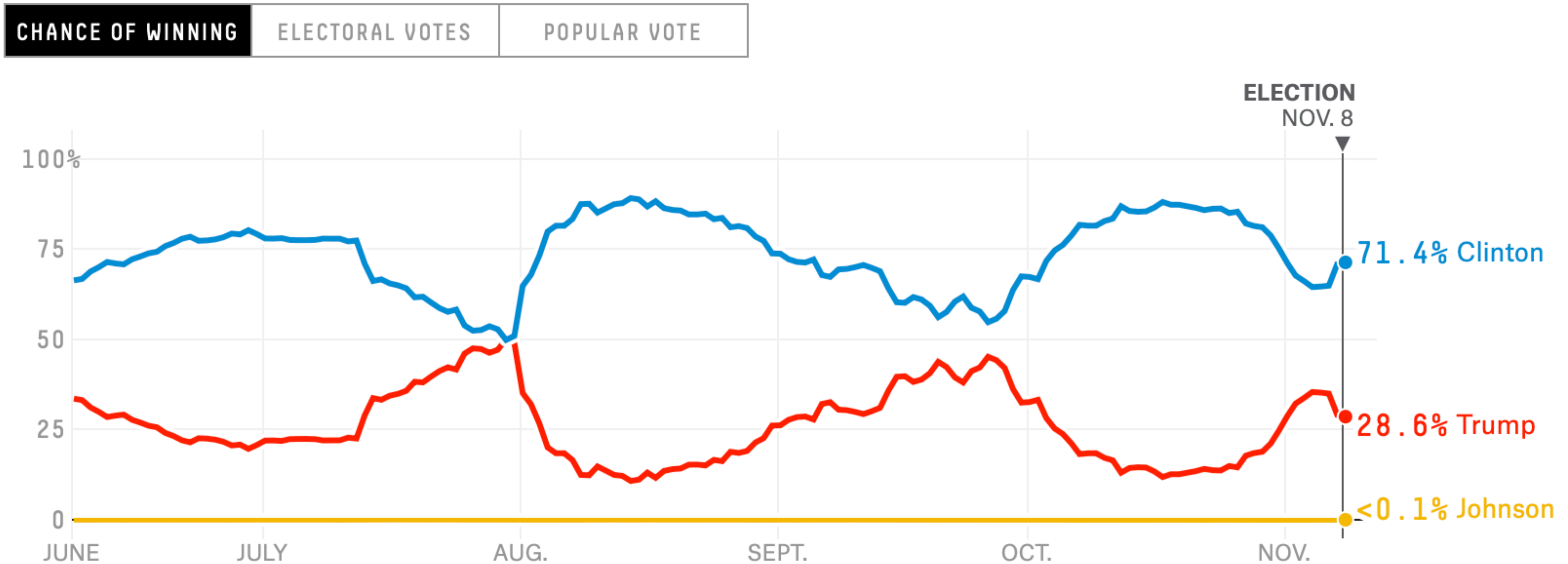
NapoleonCat.

Source: NapoleonCat.com

https://napoleoncat.com/stats/facebook-users-in-sri_lanka/2020/01

How the forecast has changed

We'll be updating our forecasts every time new data is available, every day through Nov. 8.



<https://projects.fivethirtyeight.com/2016-election-forecast/>



CASE 3

How to Predict Elections (In **Real Time**)

“Real Time” Election Model

Scientific Methods

Past Results +
Current Results Released
Data



“How many seats will
the SLPP win?”
Useful Information

Data Collection & Mining

Unstructured PDFs on <http://www.slections.gov.lk> On Past Elections

**Colombo District
Colombo-North**

PARTY/GROUP NAME	VOTES OBTAINED	PERCENTAGE
Akhila Iankai Tamil Mahasabha	17	0.03%
Our National Front	17	0.03%
United People's Freedom Alliance	10,256	16.23%
United National Party	50,571	80.03%
United People's Party	49	0.08%
Eksath Lanka Podujana Pakshaya	2	0.00%
United Socialist Party	64	0.10%
United Peace Front	4	0.01%
People's Liberation Front	1,647	2.61%
Jana Setha Peramuna	3	0.00%
Tamil United Liberation Front	162	0.26%
Nawa Sama Samaja Party	18	0.03%
Nawa Sinhala Urumaya	2	0.00%
Frontline Socialist Party	23	0.04%
Democratic Party	121	0.19%
Bochu Jana Peramuna	45	0.07%
Mauhimu Janatha Pakshaya	2	0.00%
The Liberal Party	4	0.01%
Sri Lanka National Force	0	0.00%
Sri Lanka Vanguard Party	1	0.00%
Socialist Equality Party	2	0.00%
Independent Group 01	81	0.13%
Independent Group 02	6	0.01%
Independent Group 03	1	0.00%
Independent Group 04	0	0.00%
Independent Group 05	5	0.01%
Independent Group 06	1	0.00%
Independent Group 07	1	0.00%
Independent Group 08	10	0.02%
Independent Group 09	5	0.01%
Independent Group 10	12	0.02%
Independent Group 11	29	0.05%
Independent Group 12	6	0.01%
Independent Group 13	1	0.00%
Independent Group 14	7	0.01%
Independent Group 15	18	0.03%
TOTAL VALID VOTES	63,193	93.97%
REJECTED VOTES	4,055	6.03%
TOTAL POLLED	67,248	74.91%
REGISTERED NO OF ELECTORS	89,771	



Parse PDF

Structured Data (CSV, JSON etc) On Past Elections

287 Lines (287 sloc) | 42.8 KB

Raw Blame

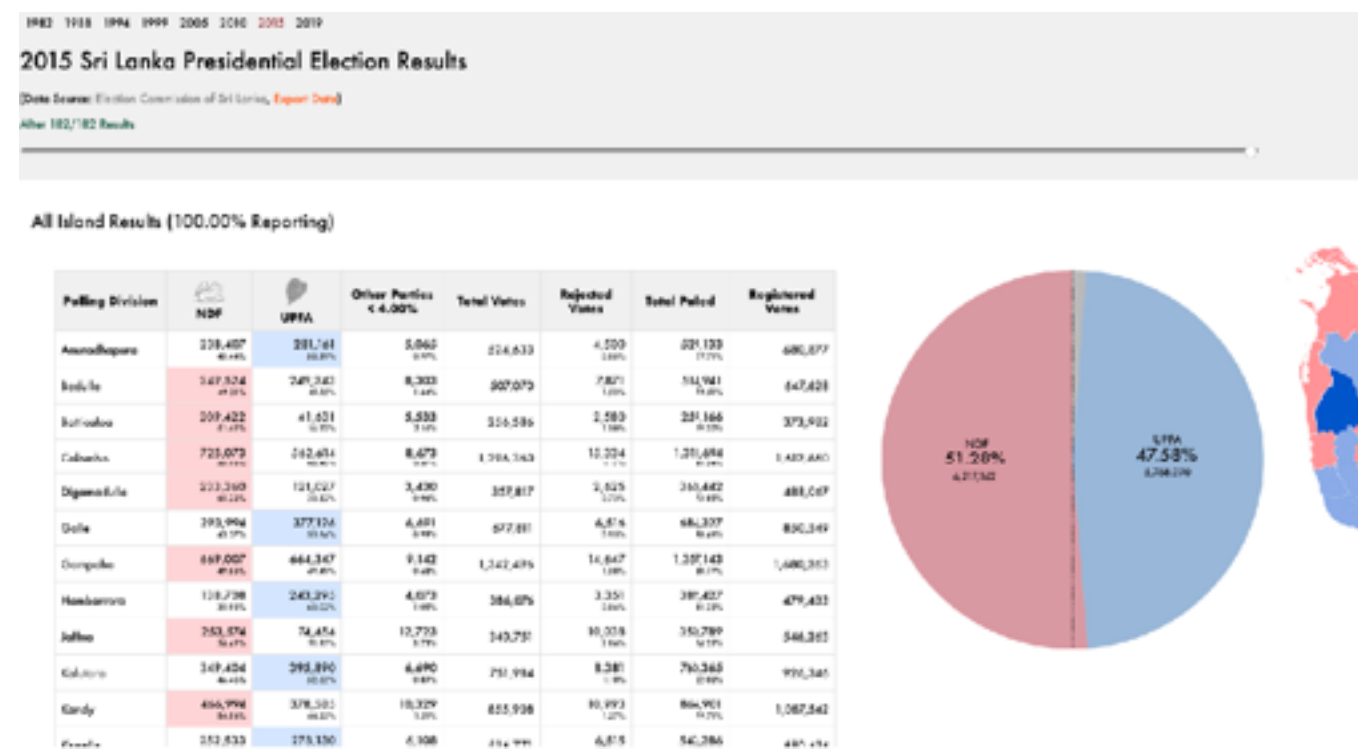
Search this file...

	type	pd_code	ed_code	pd_name	ed_name	electors	polled	rejected	valid	UNP	UPFA	PLF	ITAK	SLMC	EPDP	ACMC
2	lk	-	-	-	-	15044490	11684111	516926	11137062	5098916	4732664	544154	515963	44193	33481	33102
3	ed	-	1	-	Colombo	1586598	1252271	43372	1207610	640743	474063	81391	0	0	0	0
4	ed	-	2	-	Gampaha	1637537	1280647	56246	1223119	577004	549958	87880	0	0	0	0
5	ed	-	3	-	Kalutara	897349	719001	21366	697369	310234	338801	38475	0	0	0	0
6	ed	-	4	-	Kandy	1049160	830165	37065	792584	440761	309152	30669	0	0	0	0
7	ed	-	5	-	Matale	379675	298917	21537	277169	138241	126315	10947	0	0	0	0
8	ed	-	6	-	Nuwara-Eliya	534150	420734	32788	386957	228920	147348	5590	0	0	0	0
9	ed	-	7	-	Galle	819666	639318	15107	624027	265180	312518	37778	0	0	0	0
10	ed	-	8	-	Matara	623818	490409	12692	474180	186675	250505	35270	0	0	0	0
11	ed	-	9	-	Hambantota	462911	375885	10056	365663	130433	196980	36527	0	0	0	0

https://github.com/nuuuwan/elections_lk

JSON Endpoint

<https://results.elections.gov.lk>



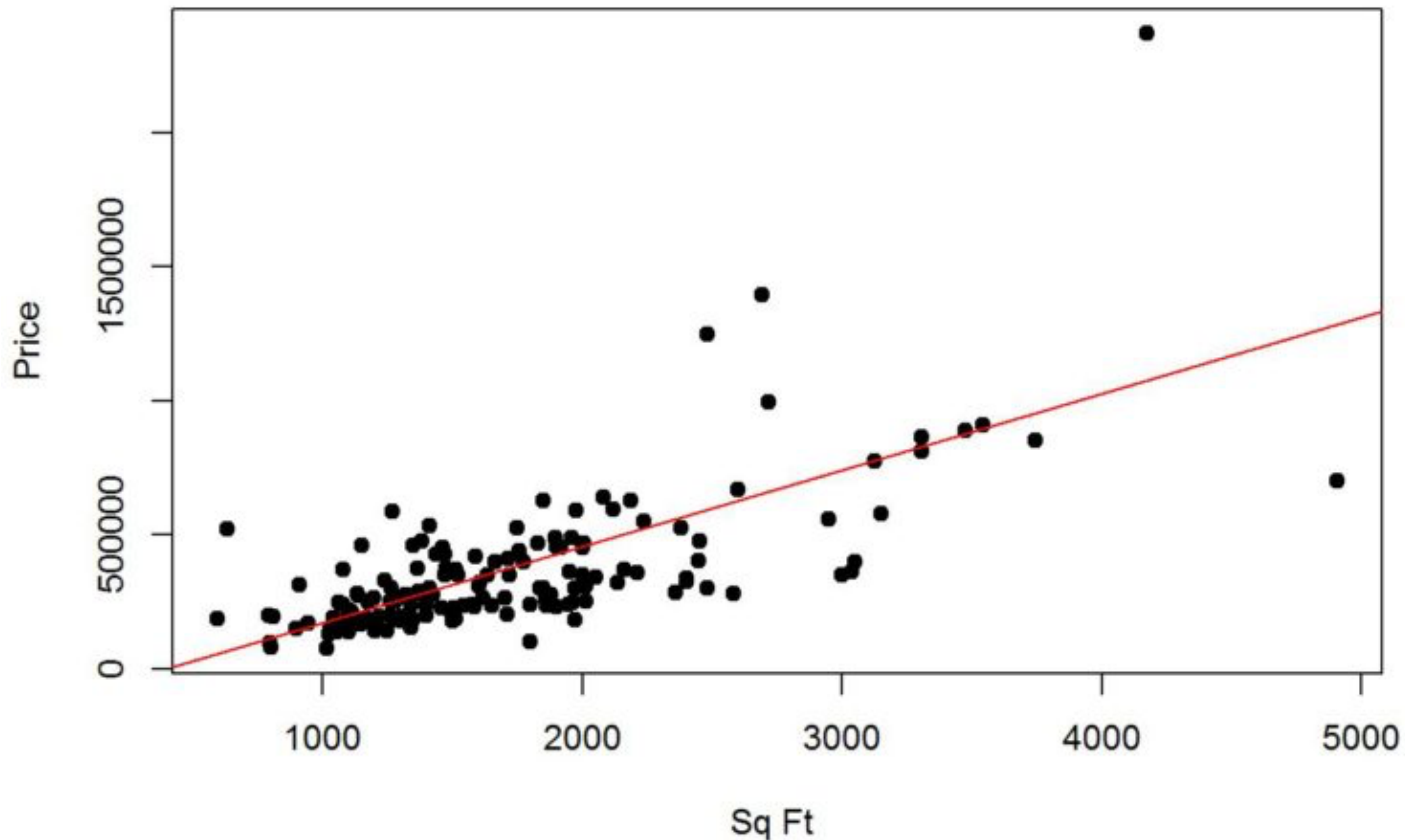
Realtime Download



<https://results.elections.gov.lk>

Machine Learning

Price and Sq Ft

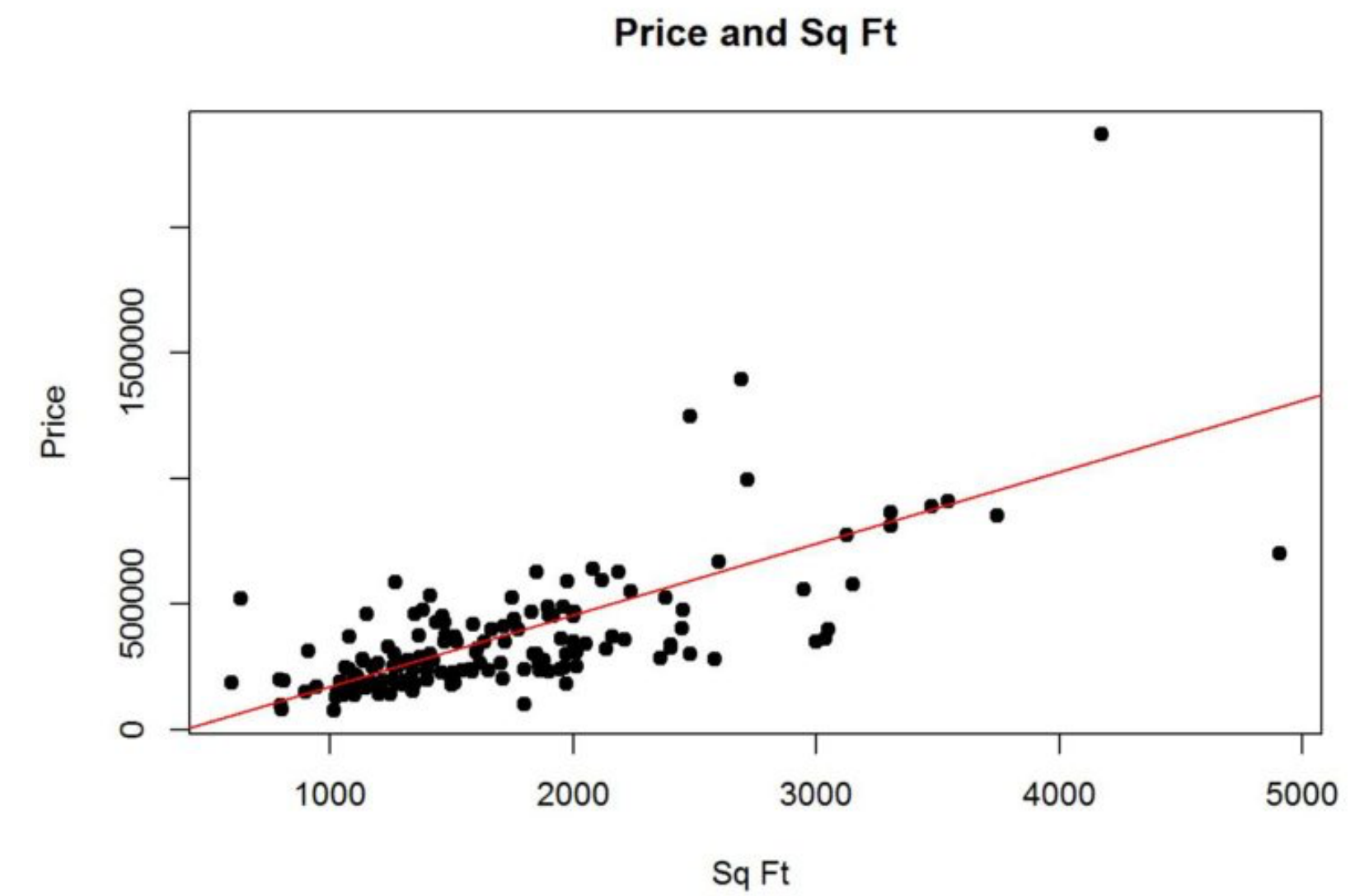


$$y = m \cdot x + c$$

$$\text{House Price} = m \cdot \text{House Area} + c + \text{Error}$$

$$\text{House Price} = f(\text{House Area}) + \text{Error}$$

$$\text{“What you want to know”} = f(\text{“What you know”}) + \text{Error}$$

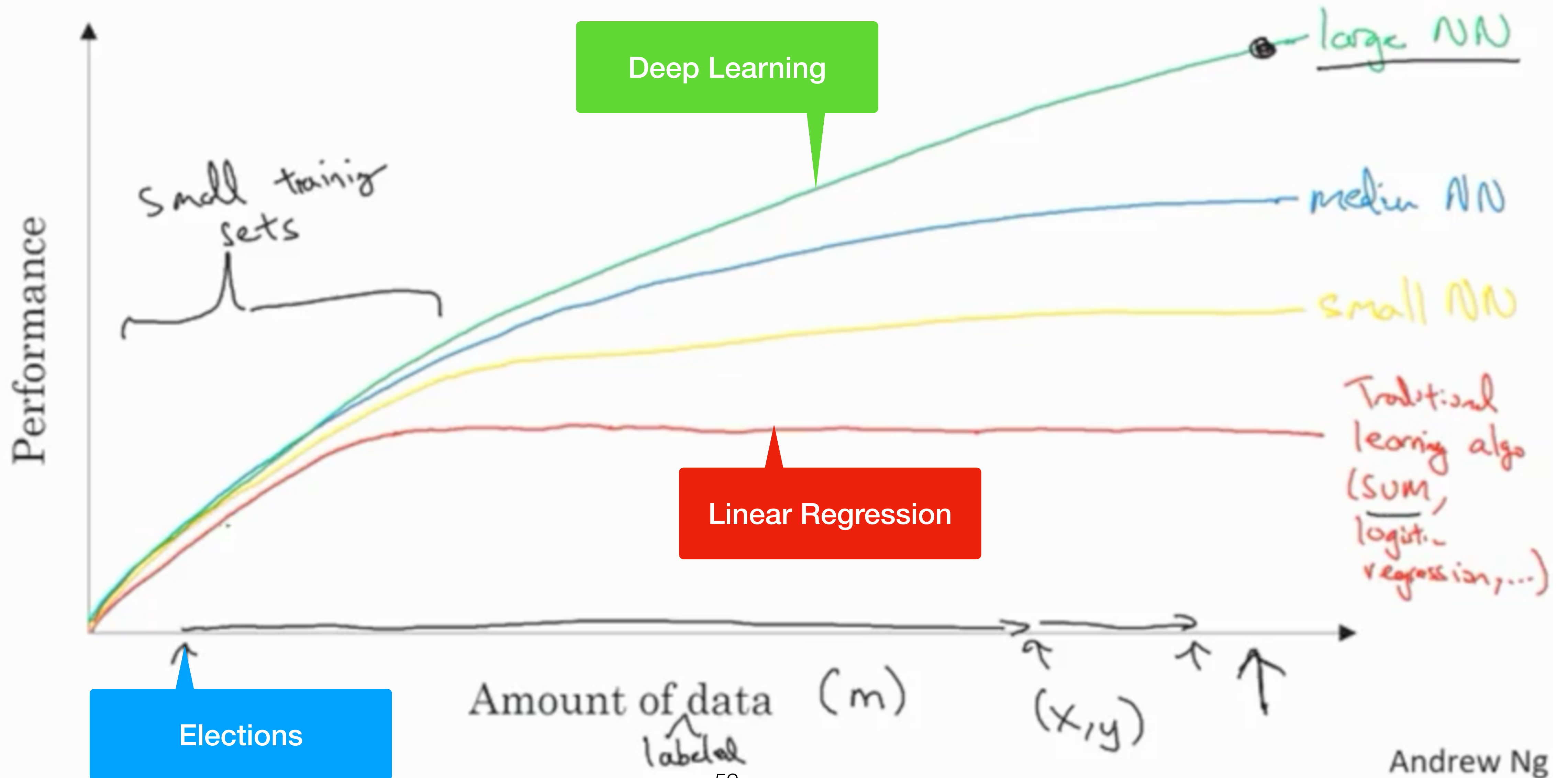



```
# given X, Y
regressor = LinearRegression()
regressor.fit(X, Y)
C = regressor.intercept_
M = list(regressor.coef_)
```

$y =$ “Seats for SLPP in the Colombo ED”

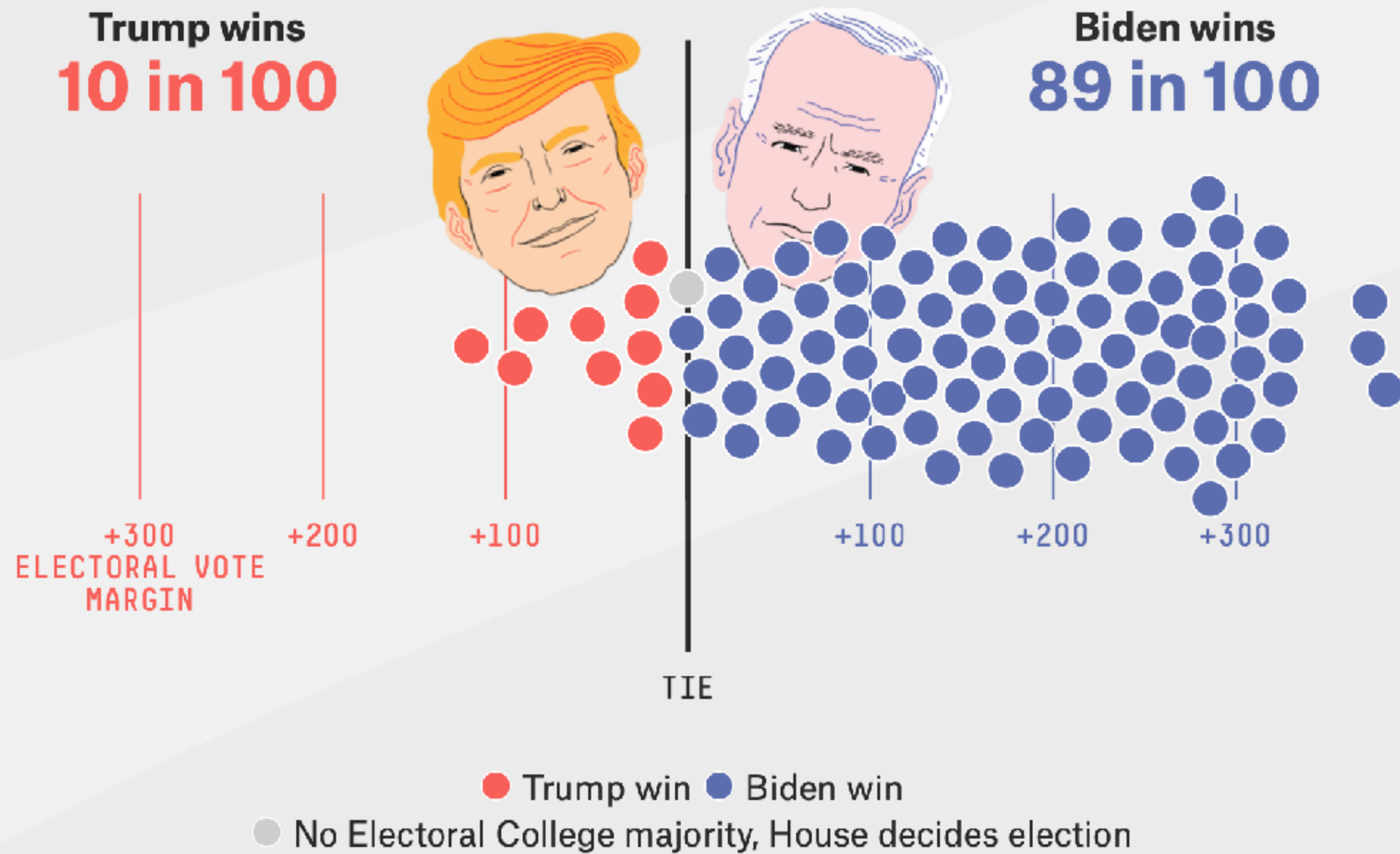
$x = [$
 “Vote % for SLPP in Habaraduwa”,
 “Vote % for SLPP in Galle”,
 ...
 $]$

Scale drives deep learning progress



Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.



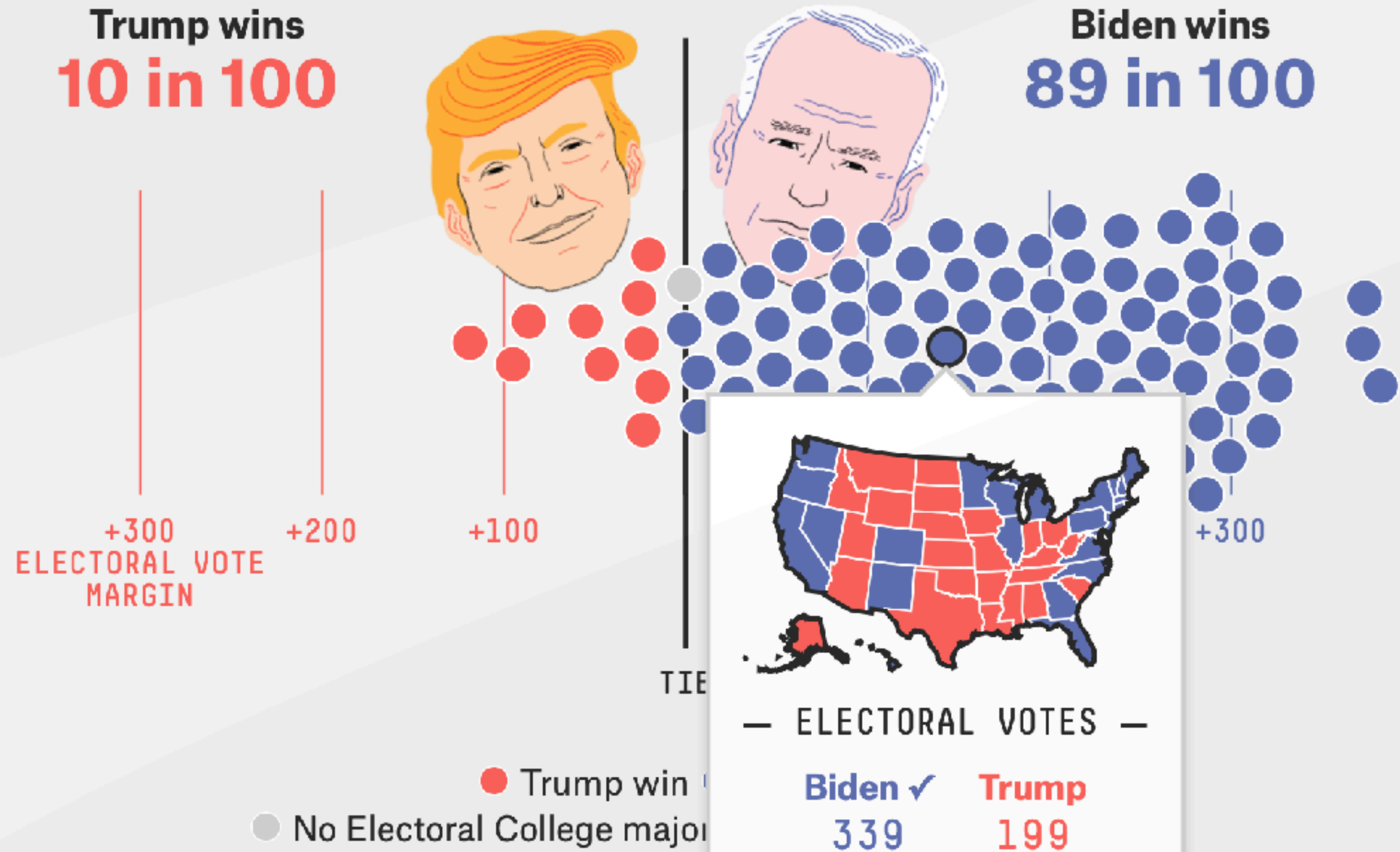
<https://projects.fivethirtyeight.com/2020-election-forecast/>

Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.

Trump wins
10 in 100

Biden wins
89 in 100



<https://projects.fivethirtyeight.com/2020-election-forecast/>

Model: “Seats for SLPP in the Colombo ED”

Model Simulation Output:

10	11	11	11	11
11	11	11	11	12
12	12	12	12	12
12	12	12	13	14

= 11 to 13 *

= 12 ± 1 *

* At 90% Confidence

Model: “Seats for SLPP in the Colombo ED”

Model Simulation Output:

10	11	11	11	11
11	11	11	11	12
12	12	12	12	12
12	12	12	13	14

* At 90% Confidence

Data Visualisation & Presentation

Results Released

129

66% Votes Reporting¹

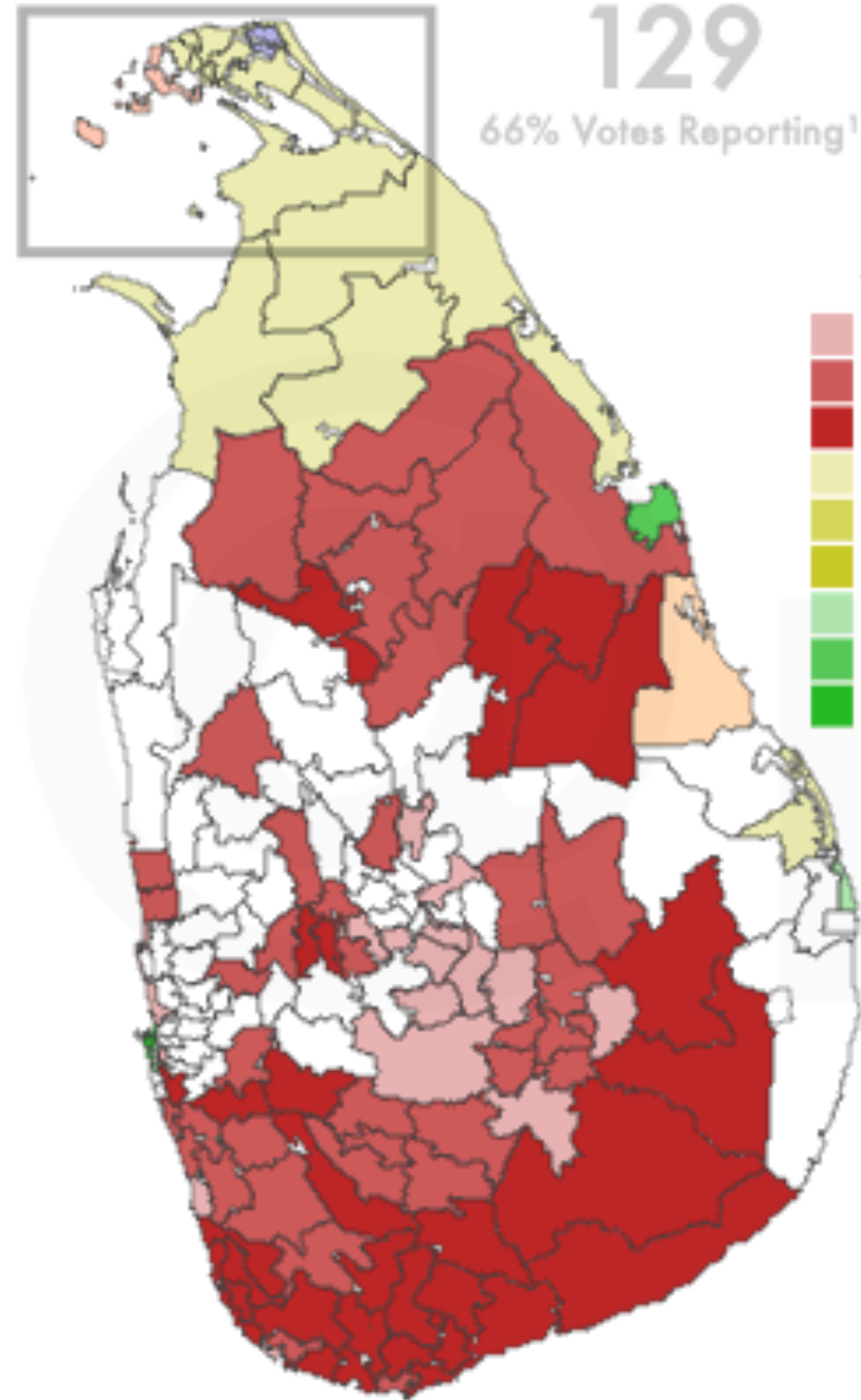
2020 Sri Lankan Parliamentary Election

Kilinochchi

Jaffna District

Projected² Seats Islandwide

SLPP majority likely



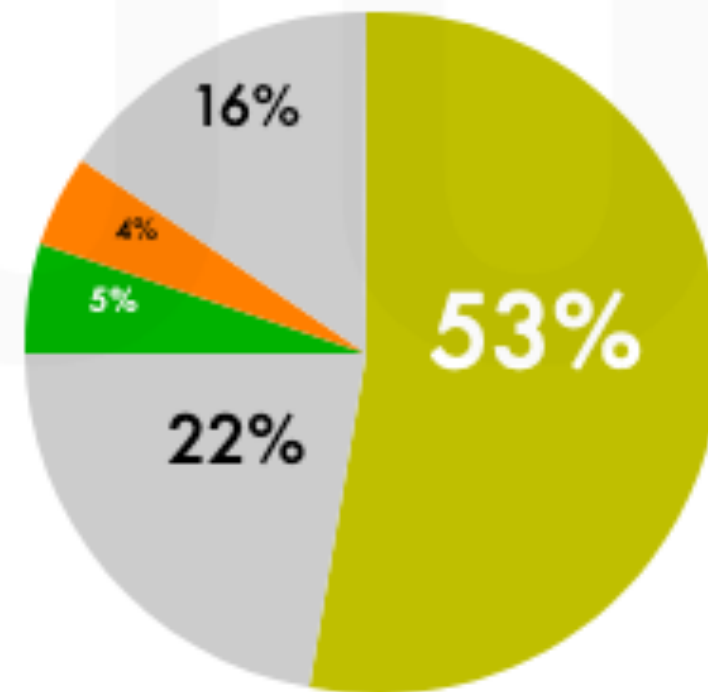
Winning Party

- SLPP < 60%
- SLPP 60% - 70%
- SLPP > 70%
- ITAK < 60%
- ITAK 60% - 70%
- ITAK > 70%
- SJB < 60%
- SJB 60% - 70%
- SJB > 70%



Kilinochchi

Turnout
74%



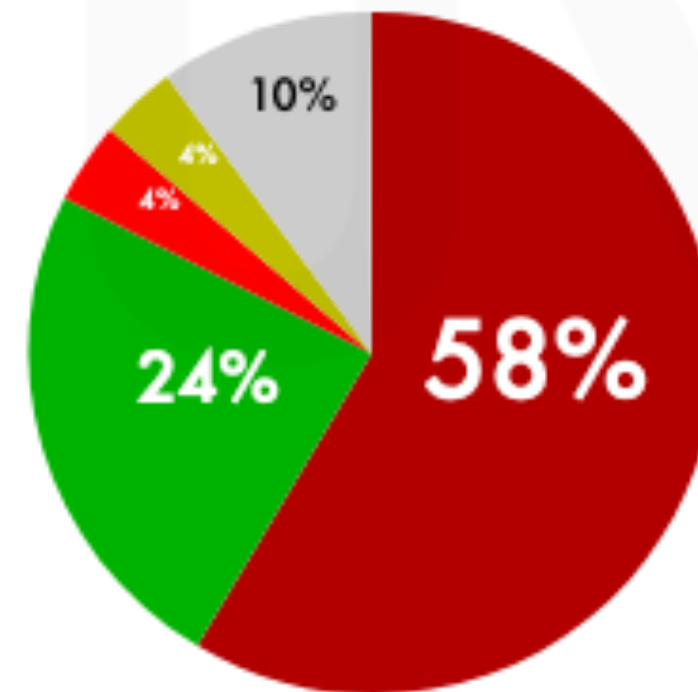
Latest Result

ITAK
31,156
IND05_D10
13,339
SJB
3,050
AITC
2,528
Others
9,268



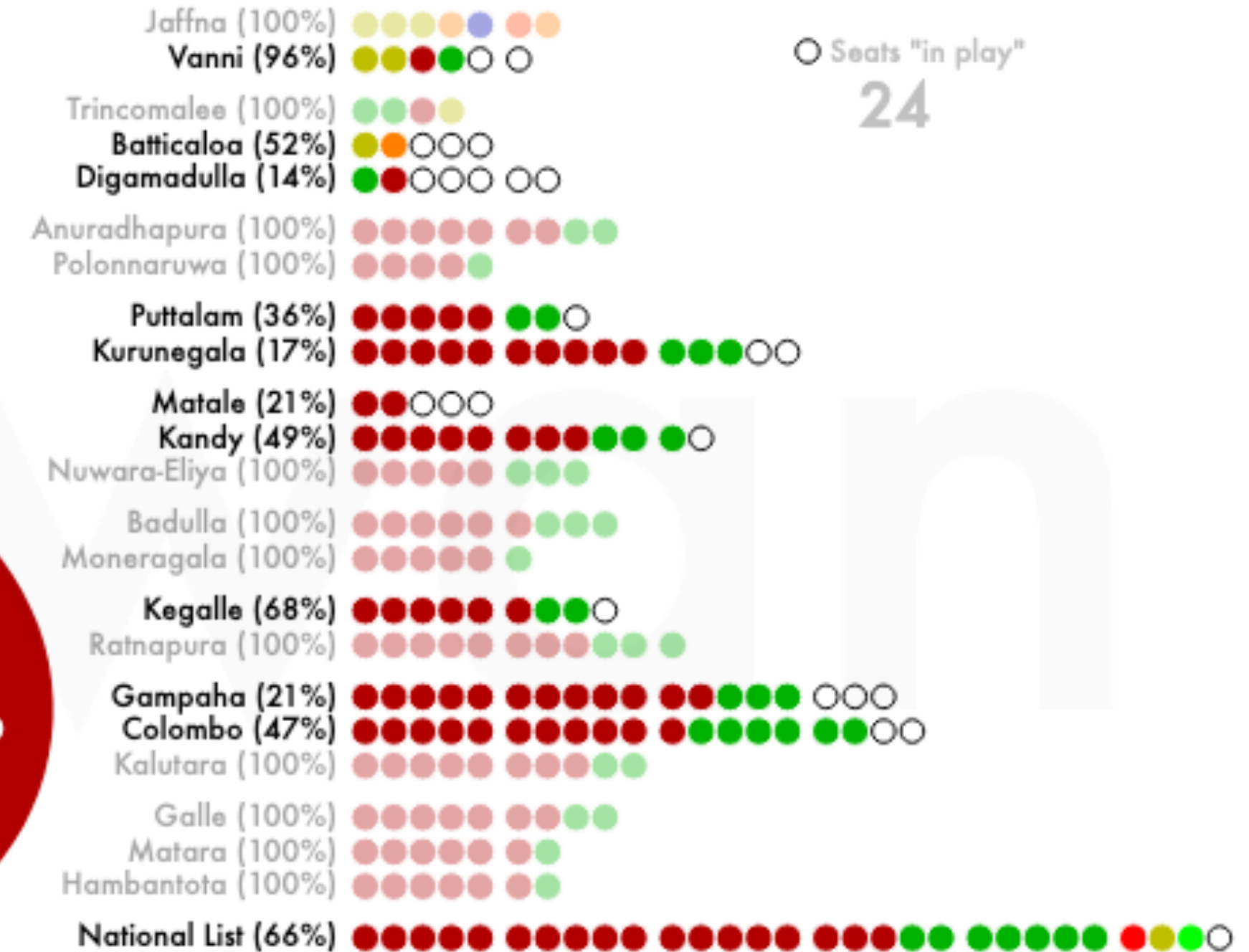
Islandwide

Turnout
76%



Cumulative Results

SLPP
4,475,758
SJB
1,839,292
JJB
294,928
ITAK
274,423
Others
769,444



Reported Seats

SLPP
63
SJB
21
ITAK
4
AITC
1
Others
3

Projected² Seats

SLPP
147 ± 10
SJB
59 ± 10
ITAK
10 ± 2
JJB
3 ± 2

¹ Based on historical data
² 95% Confidence

Polling Division - Kilinochchi

[Back](#)

Jaffna District

ITAK Ilankai Tamil Arasu Kadchi	52.5% 31,156
IND05_D10 Jafna-Independent Group 5	22.48% 13,339
SJB Samagi Jana Balawegaya	5.14% 3,050
AITC Ahila Ilankai Thamil Congress	4.26% 2,528
EPDP Eelam People's Democratic Party	3.98% 2,361
SLFP Sri Lanka Freedom Party	3.08% 1,830
TMTK Thamil Makkal Thesiya Kuttani	3.08% 1,827
IND07_D10 Jafna-Independent Group 7	0.77% 456
UNP United National Party	0.58% 346
IND09_D10 Jafna-Independent Group 9	0.45% 267

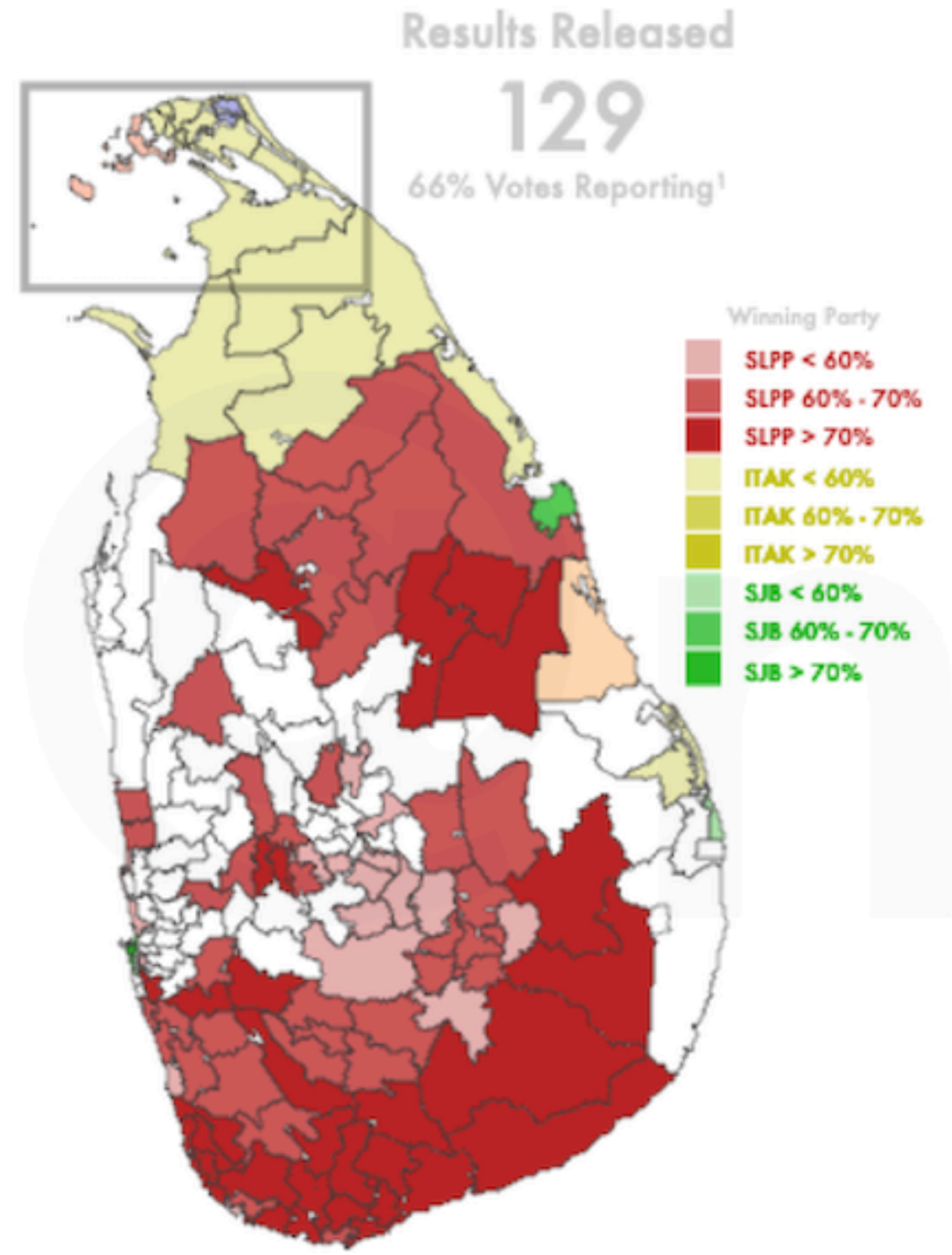
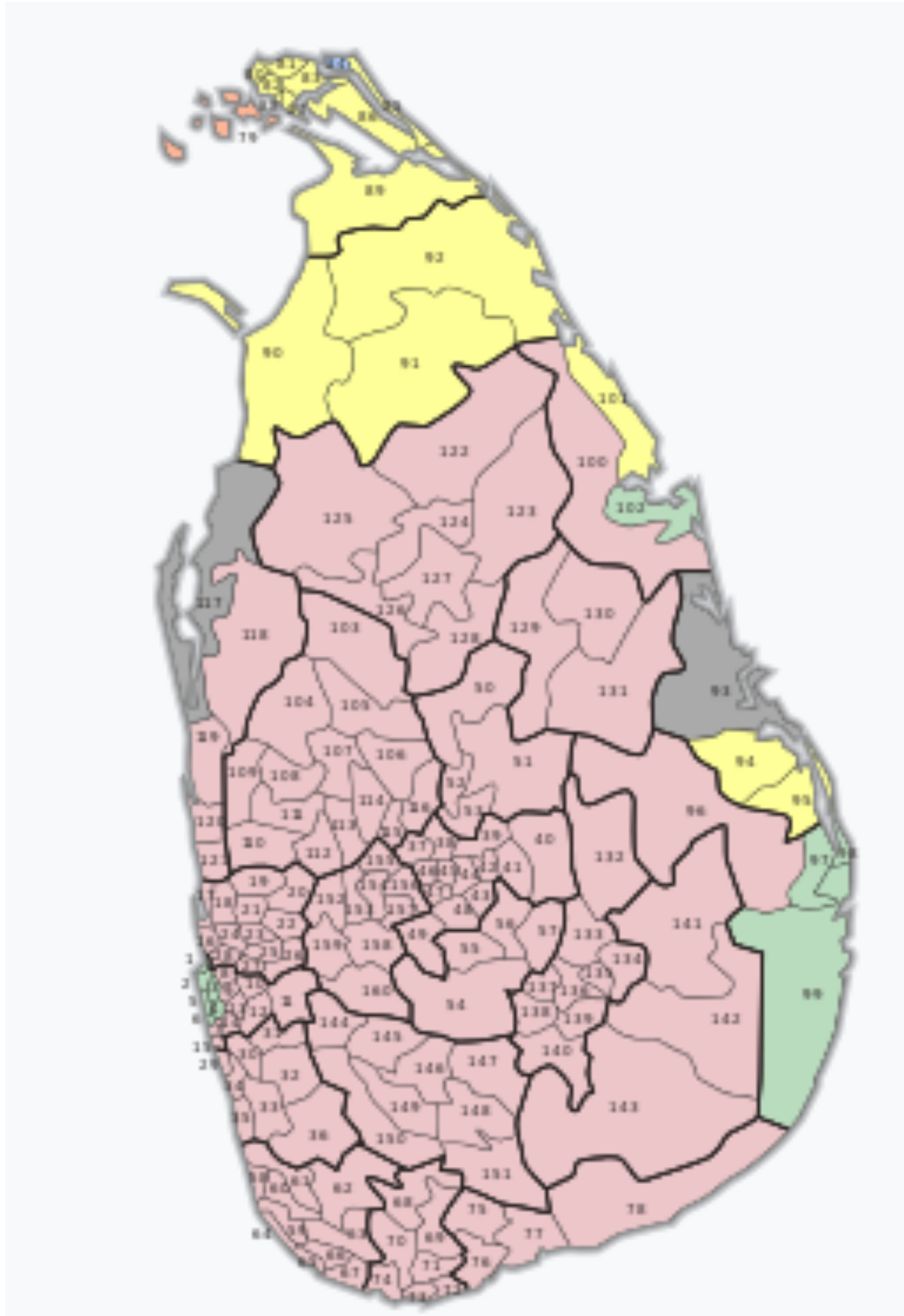
Latest Result

ITAK
31,156
IND05_D10
13,339
SJB
3,050
AITC
2,528
Others
9,268

Cumulative Results

SLPP
4,475,758
SJB
1,839,292
JJB
294,928
ITAK
274,423
Others
769,444

Minimalism



Maximalism

“Vote % for SLPP in the Colombo E.D.”

= 11 to 13

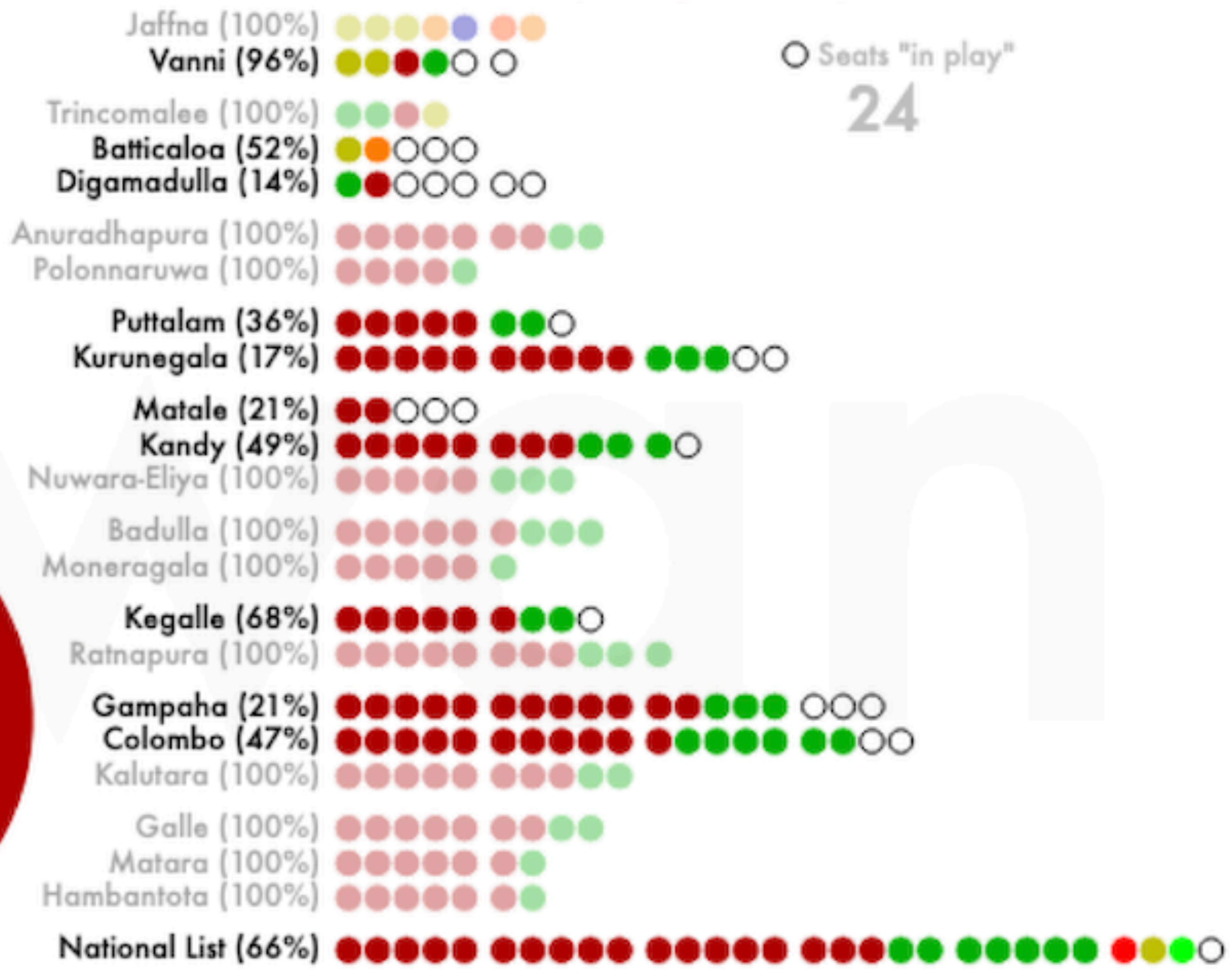
= 12 ± 1

At 90% Confidence



Projected² Seats Islandwide

SLPP majority likely



Reported Seats

SLPP

63

SJB

21

ITAK

4

AITC

1

Others

3

Projected² Seats

SLPP

147 ± 10

SJB

59 ± 10

ITAK

10 ± 2

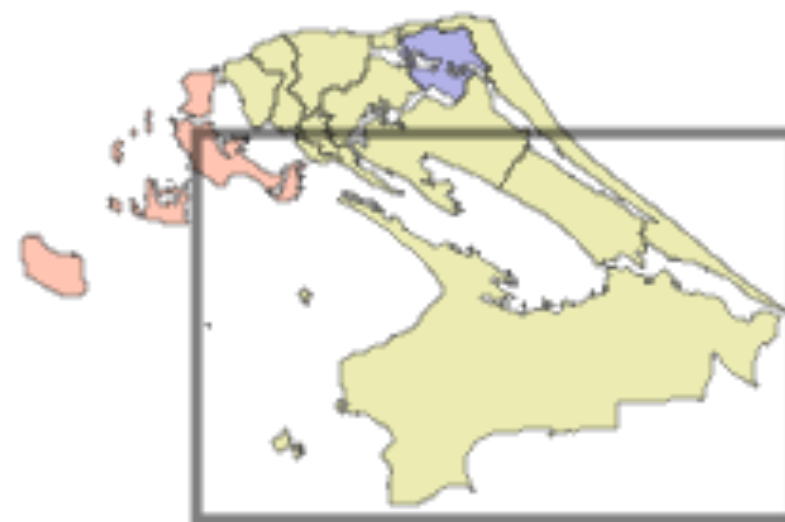
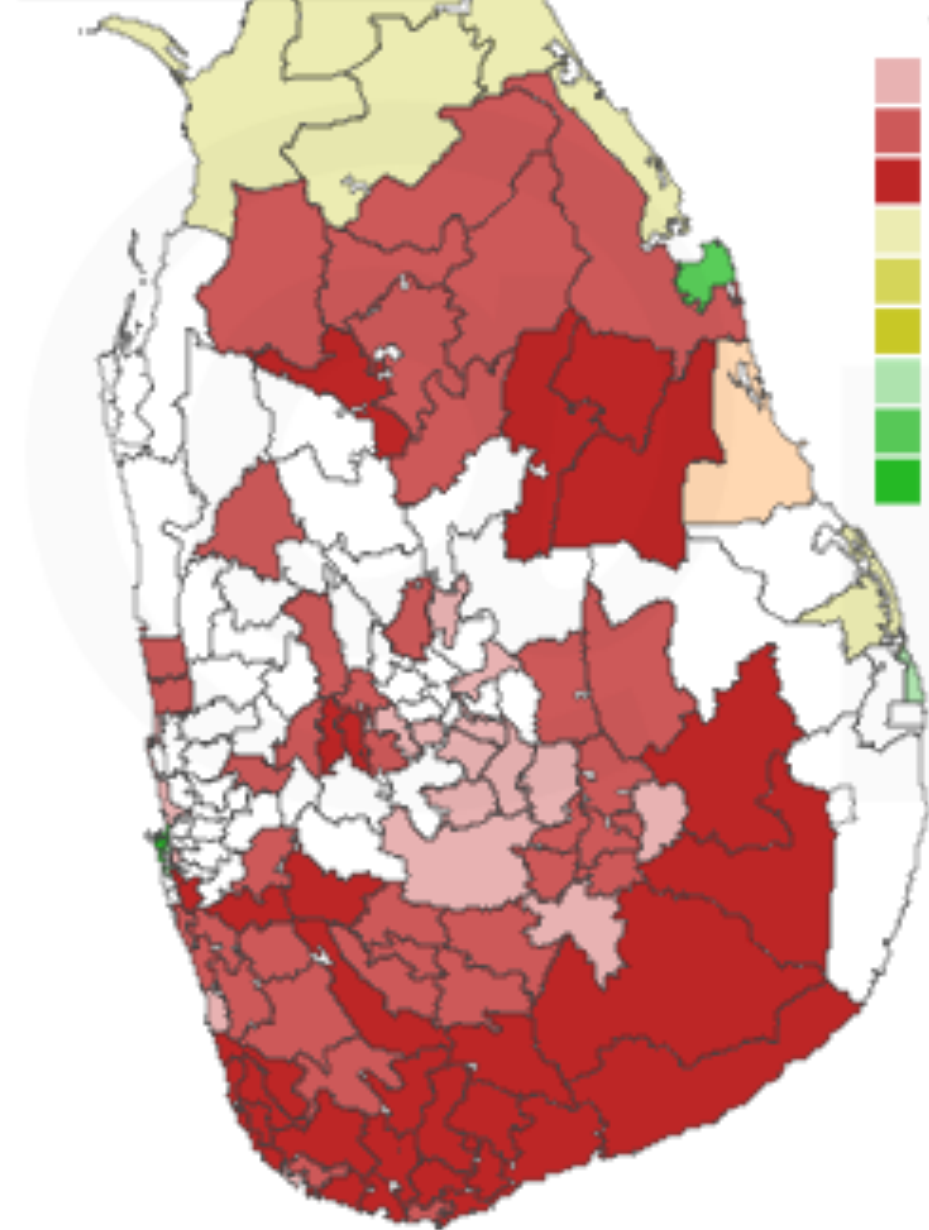
JJB

3 ± 2

Results Released

129

66% Votes Reporting¹



2020 Sri Lankan Parliamentary Election

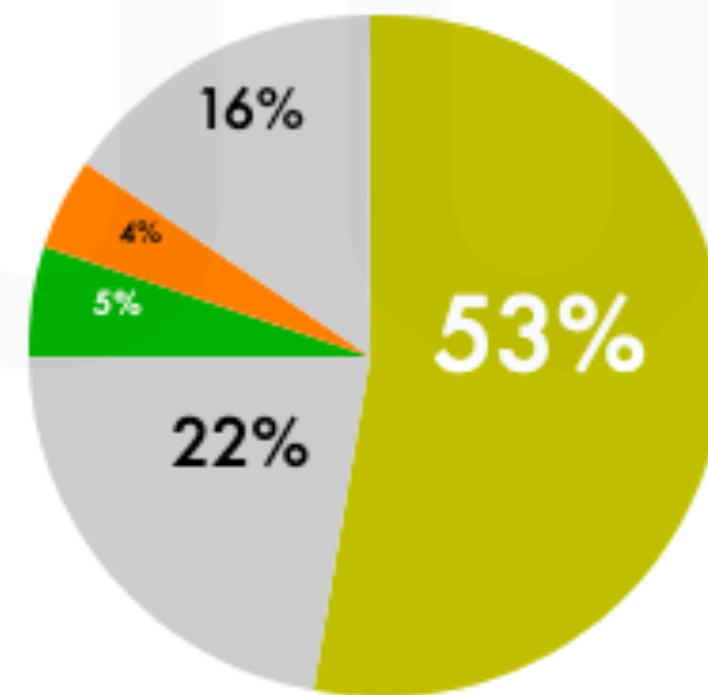
Kilinochchi

Jaffna District



Kilinochchi

Turnout
74%



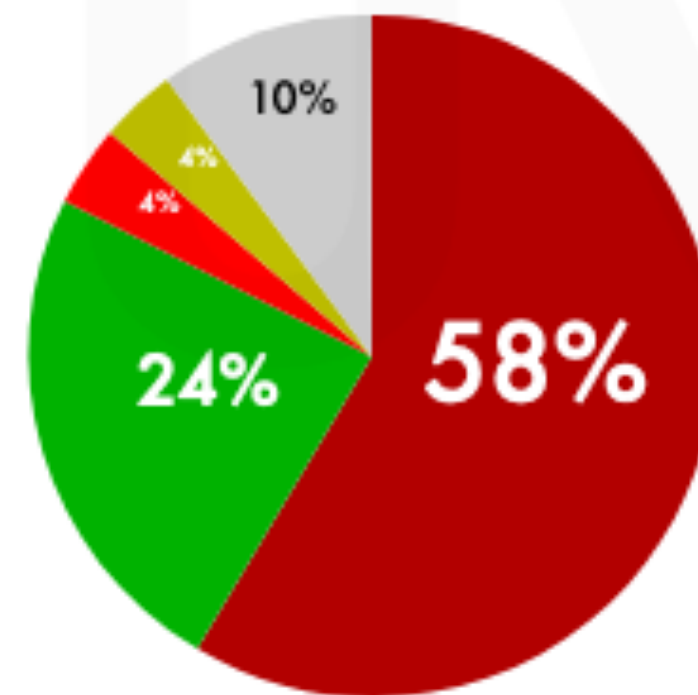
Latest Result

ITAK
31,156
IND05_D10
13,339
SJB
3,050
AITC
2,528
Others
9,268



Islandwide

Turnout
76%

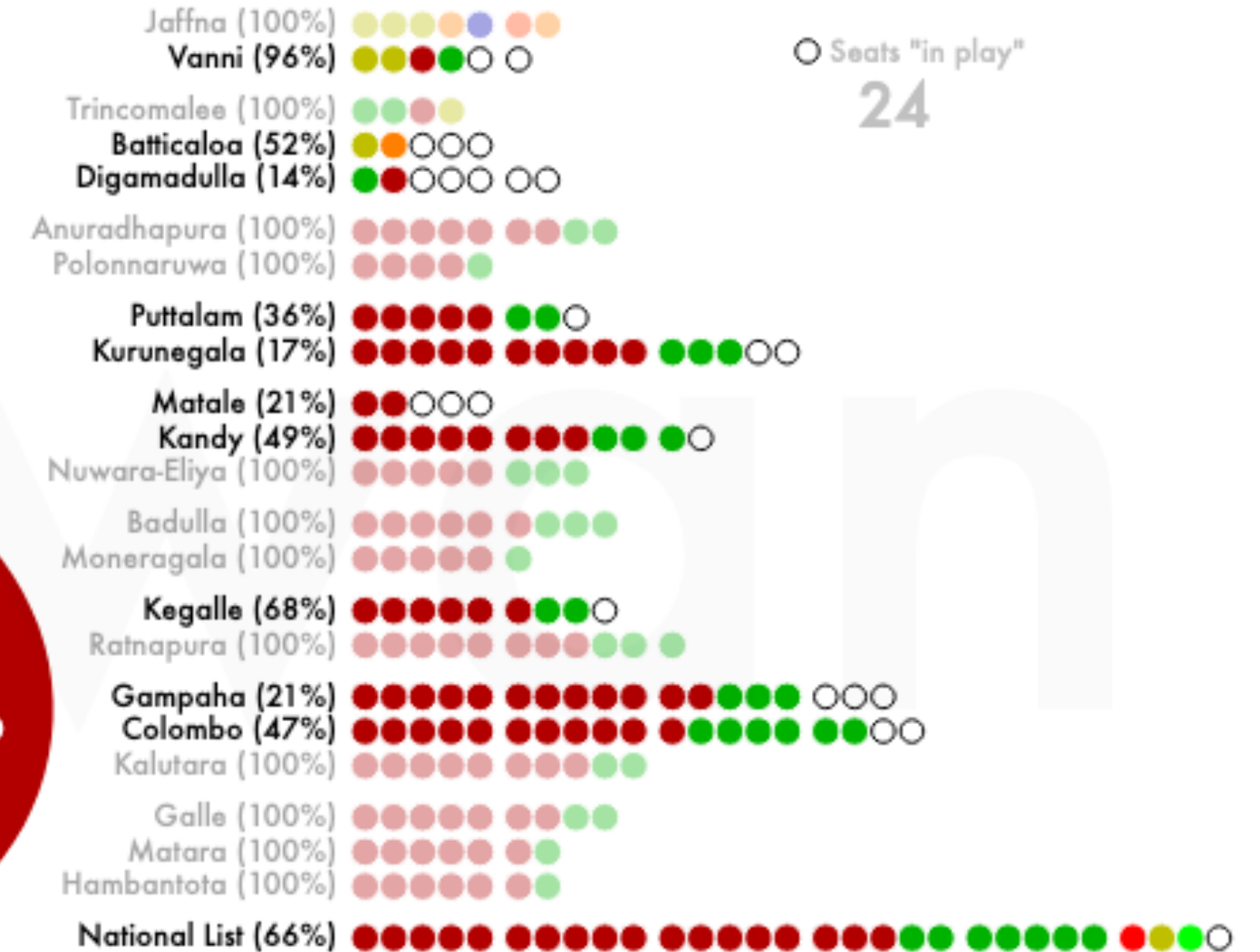


Cumulative Results

SLPP
4,475,758
SJB
1,839,292
JJB
294,928
ITAK
274,423
Others
769,444

Projected² Seats Islandwide

SLPP majority likely



Reported Seats

SLPP
63
SJB
21
ITAK
4
AITC
1
Others
3

Projected² Seats

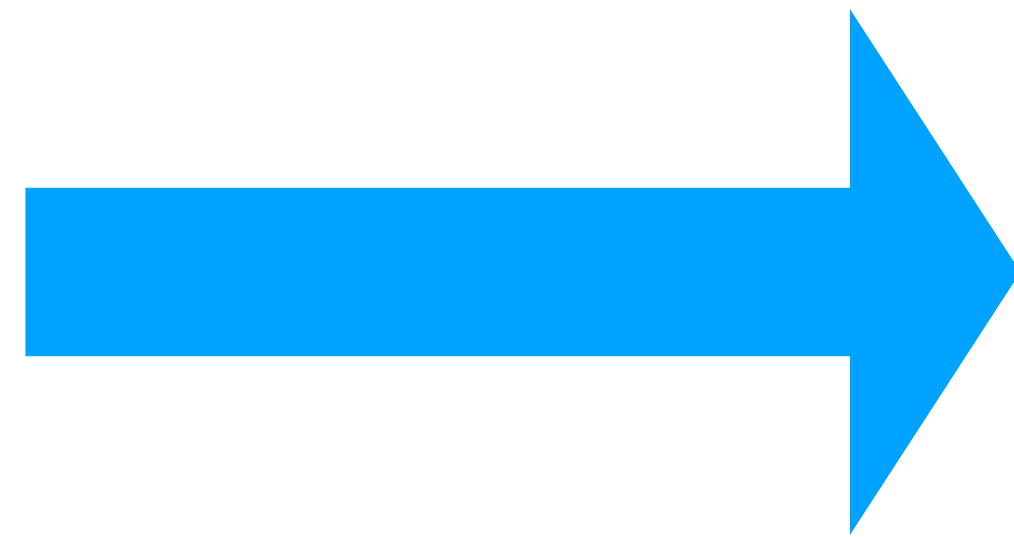
SLPP
147 ± 10
SJB
59 ± 10
ITAK
10 ± 2
JJB
3 ± 2

¹ Based on historical data
² 95% Confidence



Scientific Methods

Data

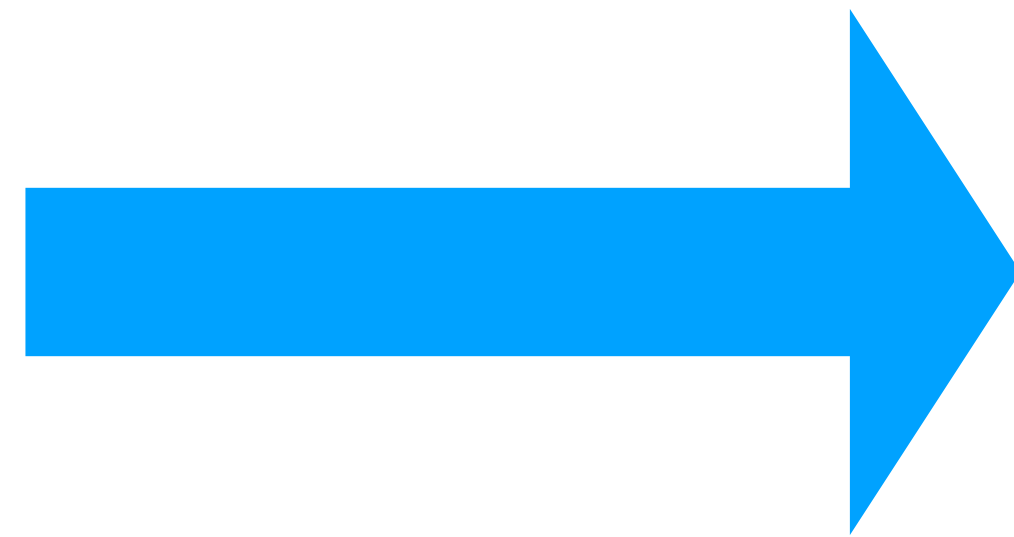


Useful
Information

Data Science

Scientific Methods (in General)

The Universe



A True
Understanding
of the Universe*

*** Even if this truth contradicts our
existing understandings and beliefs**

Pseudo-Data Science

“My Own Prejudiced
Misunderstanding of the
Universe”



Data-Driven
Disinformation

~~Data Science~~ Unethical Data Propaganda



<https://medium.com/@nuwans>

<https://twitter.com/nuuuwan>