FOUNDATIONS OF STATISTICS AND PROBABILITY FOR DATA SCIENCE

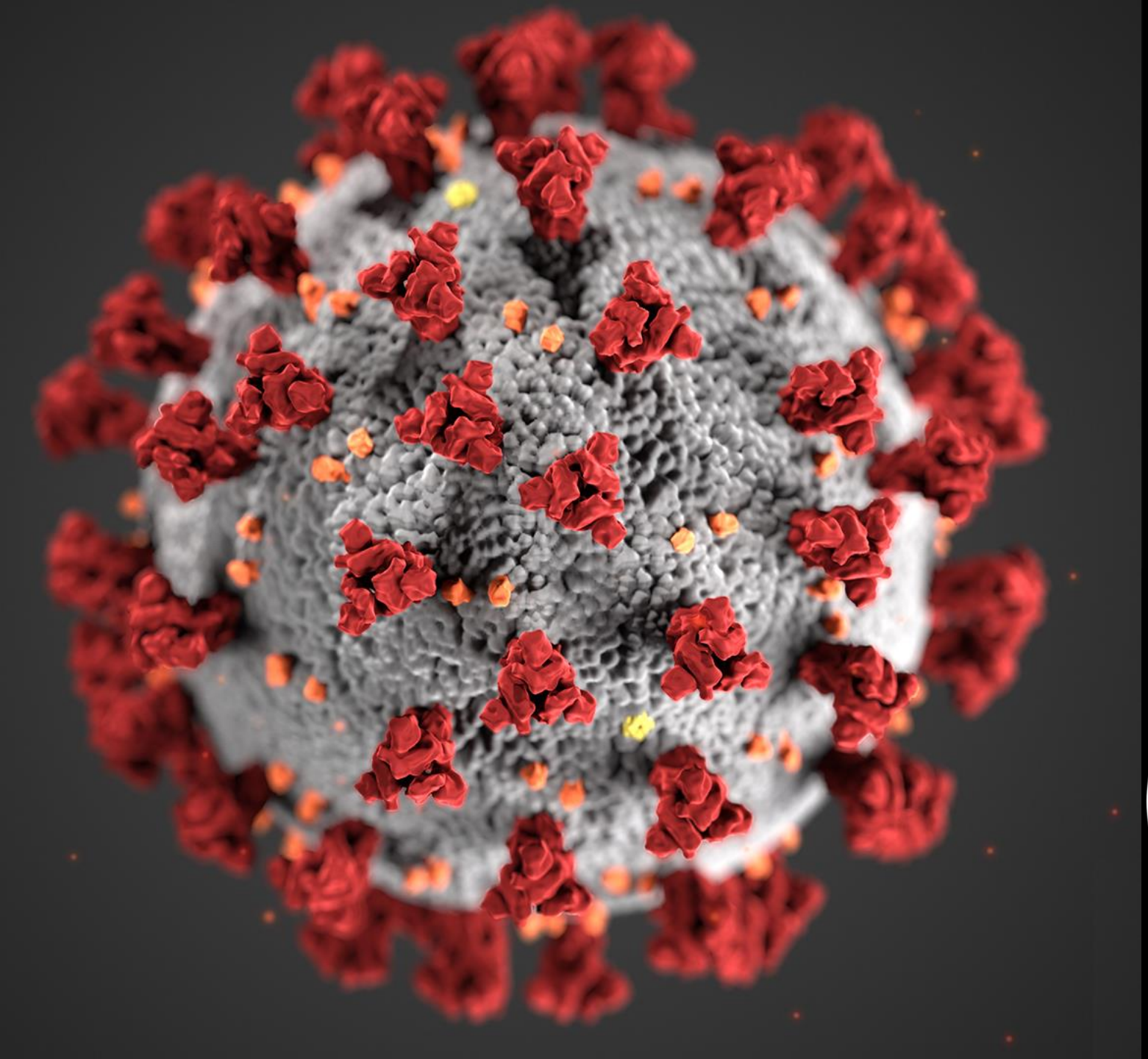
Motivation; Statistical Terminology; Descriptive Statistics: Central Tendencies, Measures of Variability; Probability Basics: Rules

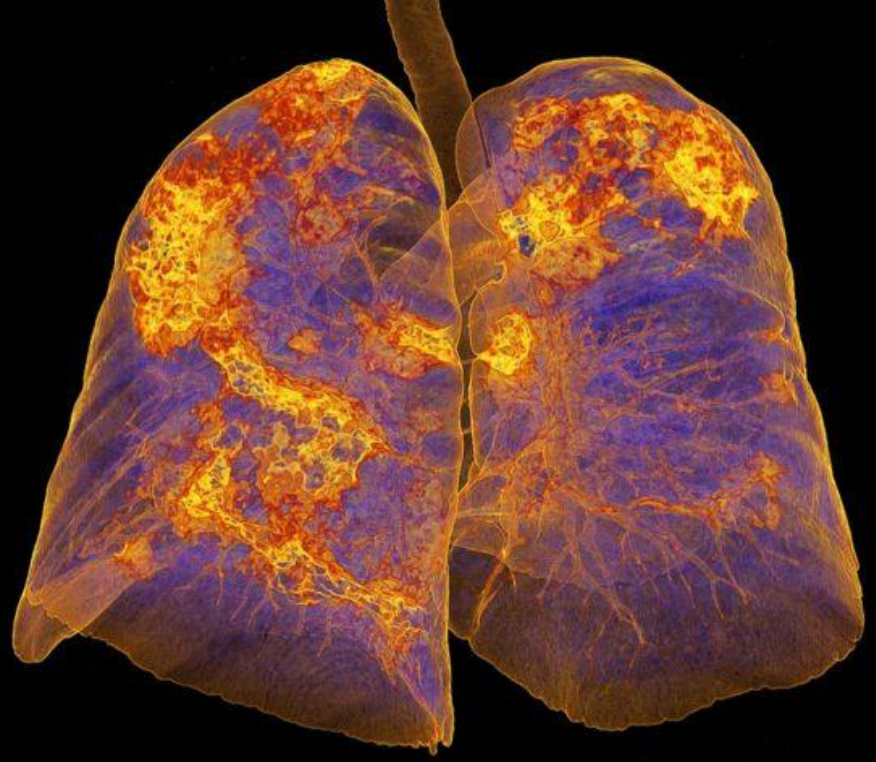
06/May/2022

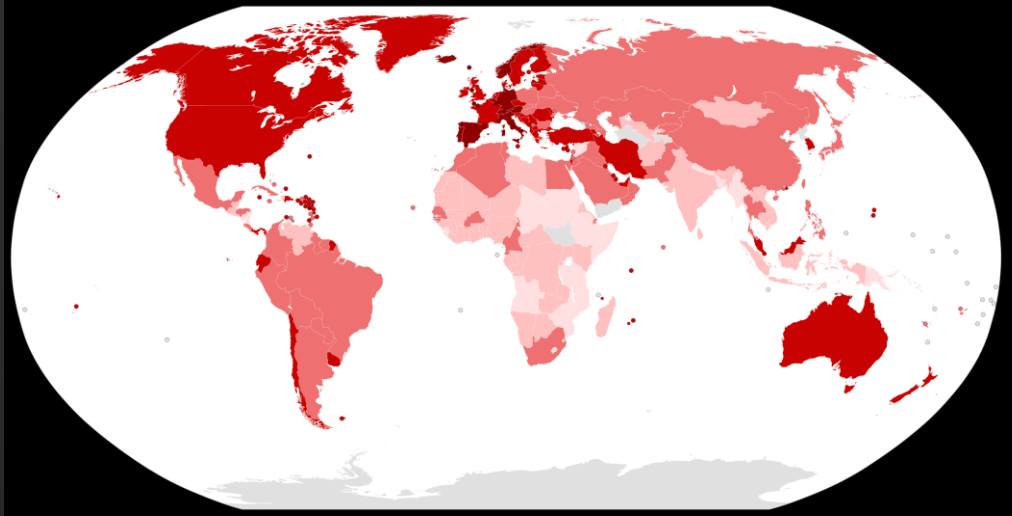
Slides prepared by Dr. Sridhar Pappu

Motivation

Statistics are all around us

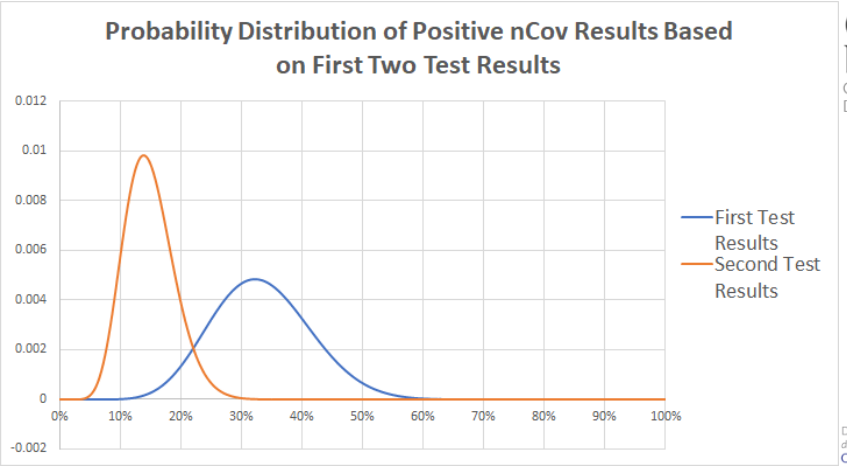


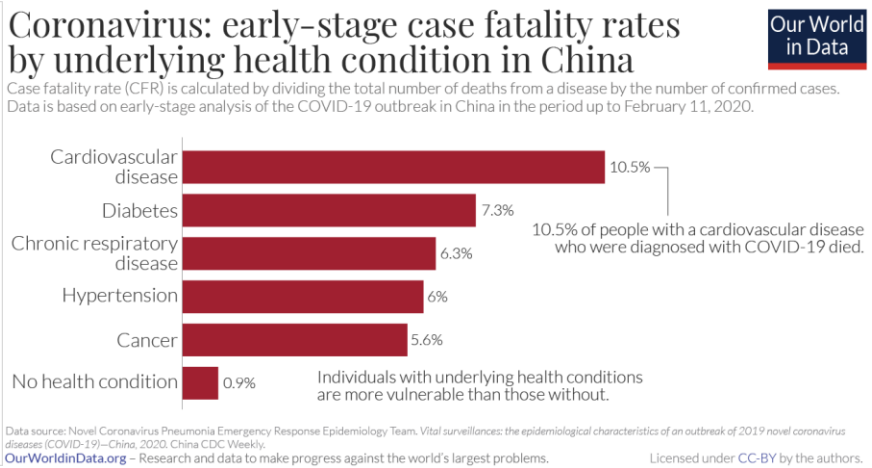


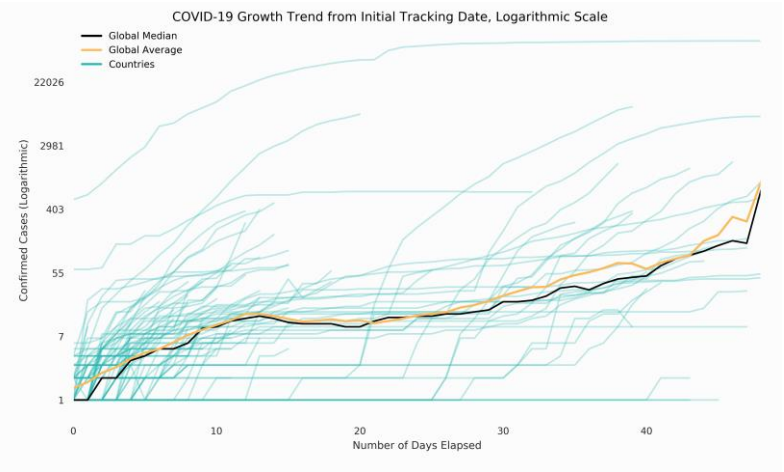


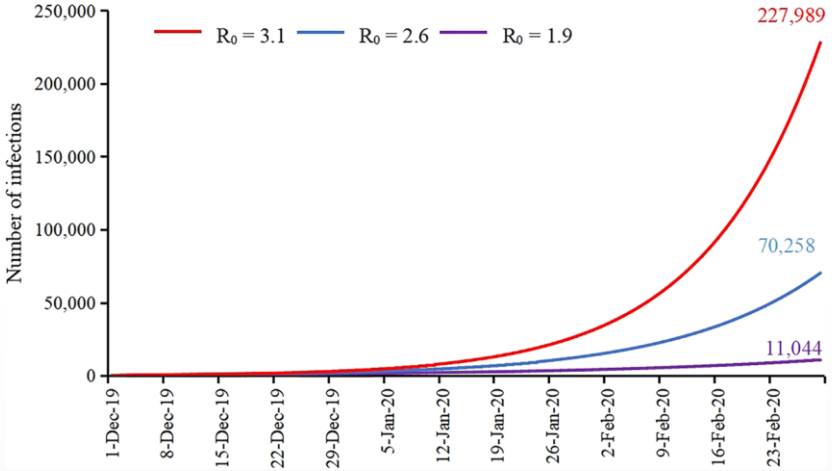
How Probability and Statistics Can Help Understand Data and the Disease Patterns

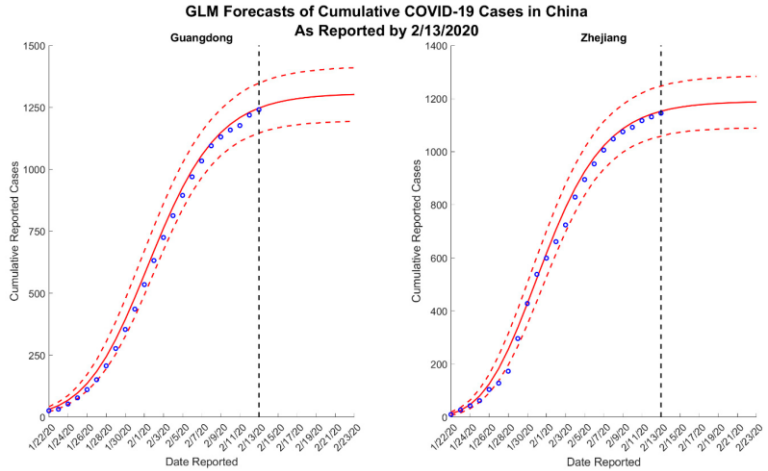
COVID-19

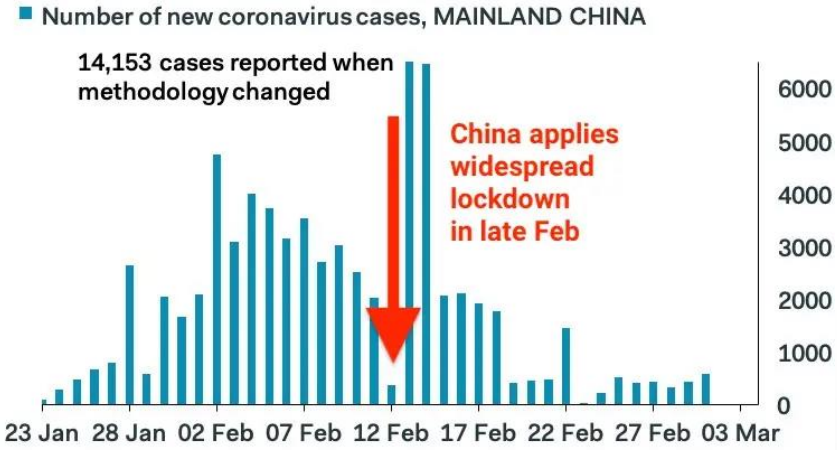

Probability Distribution of Positive nCov Results Based on First Two Test Results
First Test Results
Second Test Results

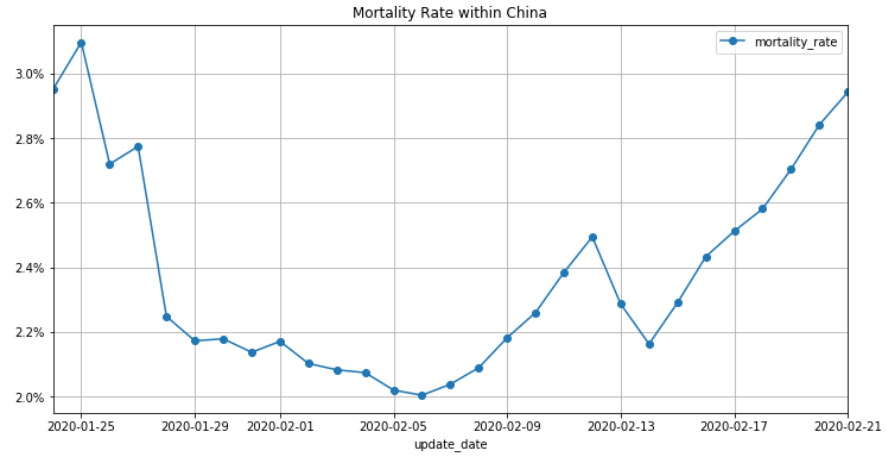

Coronavirus: early-stage case fatality rates by underlying health condition in China
Our World in Data
Case fatality rate (CFR) is calculated by dividing the total number of deaths from a disease by the number of confirmed cases. Data is based on early-stage analysis of the COVID-19 outbreak in China in the period up to February 11, 2020.
Cardiovascular disease 10.5%
Diabetes 7.3%
Chronic respiratory disease 6.3%
Hypertension 6%
Cancer 5.6%
No health condition 0.9%
10.5% of people with a cardiovascular disease who were diagnosed with COVID-19 died.
Individuals with underlying health conditions are more vulnerable than those without.
Data source: Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. China CDC Weekly.
OurWorldinData.org – Research and data to make progress against the world's largest problems.
Licensed under CC-BY by the authors.

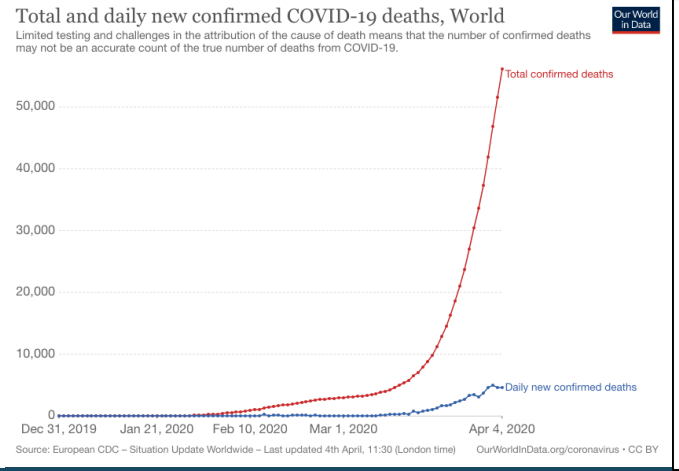

COVID-19 Growth Trend from Initial Tracking Date, Logarithmic Scale
Global Median
Global Average
Countries
Confirmed Cases (Logarithmic)
Number of Days Elapsed

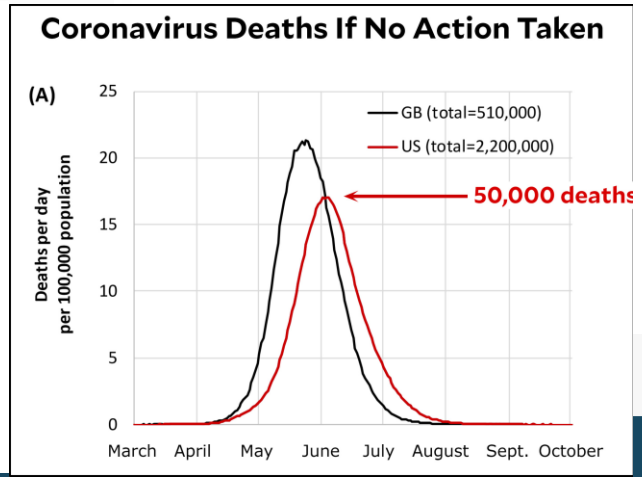

R₀ = 3.1
R₀ = 2.6
R₀ = 1.9
227,989
70,258
11,044
Number of infections

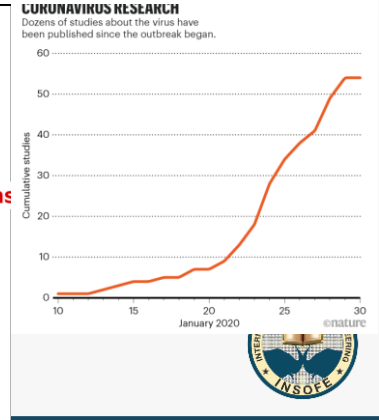

GLM Forecasts of Cumulative COVID-19 Cases in China As Reported by 2/13/2020
Guangdong
Zhejiang
Cumulative Reported Cases
Date Reported


Number of new coronavirus cases, MAINLAND CHINA
14,153 cases reported when methodology changed
China applies widespread lockdown in late Feb
23 Jan
28 Jan
02 Feb
07 Feb
12 Feb
17 Feb
22 Feb
27 Feb
03 Mar


Mortality Rate within China
mortality_rate
update_date


Total and daily new confirmed COVID-19 deaths, World
Our World in Data
Limited testing and challenges in the attribution of the cause of death means that the number of confirmed deaths may not be an accurate count of the true number of deaths from COVID-19.
Total confirmed deaths
Daily new confirmed deaths
Dec 31, 2019
Jan 21, 2020
Feb 10, 2020
Mar 1, 2020
Apr 4, 2020
Source: European CDC – Situation Update Worldwide – Last updated 4th April, 11:30 (London time)
OurWorldInData.org/coronavirus • CC BY


Coronavirus Deaths If No Action Taken
(A)
GB (total=510,000)
US (total=2,200,000)
50,000 deaths
Deaths per day per 100,000 population
March
April
May
June
July
August
Sept.
October


CORONAVIRUS RESEARCH
Dozens of studies about the virus have been published since the outbreak began.
Cumulative studies
January 2020
nature

Vax 100% effective in kids aged 12-15 yrs, say Pfizer & BioNTech

Adolescents Experienced No Serious Side Effects In Trial, Say Cos

Pfizer and BioNTech's Covid-19 vaccine is safe, effective and produces robust antibody responses in 12 to 15-year olds, they said on Wednesday, paving the way for them to seek US and European approval in weeks to use the shot in this age group.

The read-out from a clinical trial, which puts the pair ahead of other vaccine developers in the quest to protect children, will likely allow for the use of the vaccine in that group before the next school year, Albert Bourla, Pfizer's CEO, said in a statement.

Whether Covid-19 vaccines work and are safe to use on children is one of the big questions drugmakers are trying to answer. Inoculating children and young people is considered a critical step towards reaching "herd immunity" and taming the pandemic.

The Pfizer and BioNTech vaccine is already authorised for use in people starting at age 16. The new study offers the first evidence of how the vaccine will also work in school-age adolescents. In the trial of 2,260 adolescents aged 12 to 15, there were 18 cases of Covid-19 in the group that got a placebo shot and none in the group that got the vaccine, resulting in 100% efficacy in preventing Covid-19, the compa

THE TIMES OF INDIA

HYDERABAD, FRIDAY, APRIL 2, 2021

Covid virus may be getting more infectious, less lethal

Atul.Thakur@timesgroup.com

An analysis of Covid-19 cases and deaths in the ongoing second wave in India suggests that the virus maybe getting more infectious but less lethal – a course that many experts and epidemiologists had earlier anticipated. The analysis by **TOI** compared the ongoing wave with the earlier phase of outbreak in 23 large states and UTs that are witnessing a steady increase in cases.

Based on the growth rate of cases, the pandemic can be divided into two broad phases — the early phase that started in January last year, peaked around September and was cooling off until mid-February this year and the latest phase that started after February in most states (except Maharashtra and Punjab) where infections are again rising. Of course, the rise may also be due to reduced curbs and increased laxity. Still, data suggest that in five states/UTs — Maharashtra, Punjab, Gujarat, Madhya Pradesh and Chandigarh — the second wave seems to be in full flow with average daily fresh cases surpassing those in the earlier phase.

In Maharashtra, the first wave started on March 9, 2020 and ended on December 22, 2020, when daily average of cases reached its minimum value. In this 288-day period, 19 lakh cases were reported, averaging 6,606 a day. Since December 22, the state is witnessing another surge in cases and the 99 days since then till March 31, 2021 have seen 9.1 lakh cases, an average of 9,197 cases a day.

Similarly, the daily average of cases in Punjab is 1,069 for this phase compared to 532 earlier.

▶Continued on P7

CFR DOWN IN WORST-HIT AREAS

State/UT	Previous Phase			Ongoing Phase			Drop In CF Rate
	Daily Cases	Daily Deaths	CF Rate	Daily Cases	Daily Deaths	CF Rate	
Maha	6,606	170	2.6	9,197	58	0.6	75
Punjab	532	17	3.2	1,069	20	1.9	41
Gujarat	805	13	1.7	888	2	0.3	83
MP	780	12	1.5	828	3	0.4	72
Chandi.	64	1	1.6	128	1	0.6	65
India	28,784	411	1.4	27,922	156	0.6	61

Source: covid19india.org

2nd wave may peak around Apr 15-20: Study

Scientists working on a mathematical model to examine the Covid-19 spread forecast that the current wave will peak between April 15 and 20 and the number of active symptomatic cases will hit a high about a week later at the same level of around 10 lakh cases India had seen in September last year, reports **Vishwa Mohan**. The peak is likely to see between 80,000 and 90,000 new infections per day, Manindra Agrawal of IIT-Kanpur, who was involved with the national 'super model' initiative, said. The steep rise (during April 15-20) will be followed by a sharp fall over next 15-20 days, he said. P9

▶Above 45 account for 88% of deaths, P9

Cases up, but impact unchanged: Data

Sushmi.Dey@timesgroup.com

New Delhi: Even as Covid-19 cases are again rising across the country, the impact of the disease and the risk of death due to the infection on different age groups and gender remain unchanged since December, official data showed.

People above 60 years of age continue to be most vulnerable with 55% mortality reported amonunchanged sicng them, even though they account for 14% of positive cases, the latest analysis of age-wise d fected cases and mort Covid-19 showed.

Death share is among those betw years, contributing 33

88% OF DEATHS AMONG 45+

(As on April 1, 2021)

Age	% of total Covid-19 Positive cases	% of total Covid-19 deaths
<17 YEARS	8	1
18-25	13	1
26-44	39	10
45-60	26	33
>60	14	55
Gender		
MALE	63	70
FEMALE	37	30

the risk of death.

The incidence of the infection is still high in the age band of 18-44 years with 52% of positive cases. However, only 11% of total deaths are in this age bracket. Experts say the high number of cases is possibly because this is the working age population that is more likely to go out and, therefore, are likely spreaders too.

The share of positive cases is higher among men compared to women. Men accounted for 63% of total cases and Covid-19.

aharashtra, Chhattis- ccount for al 5.84 lakh

2nd wave may peak by April 20, say scientists

Vishwa.Mohan@timesgroup.com

New Delhi: Scientists working on a mathematical model to examine the Covid-19 spread forecast that the current second wave will peak somewhere between April 15-20 and the number of 'active' symptomatic cases would hit a high roughly a week later at the same level of around 10 lakh cases the country had seen in September last year.

Manindra Agrawal of IIT Kanpur, who was involved with the national 'super model' initiative, told **TOI** on Thursday the peak is expected to see between 80,000-90,000 new infections per day. "It may be off by 10-15%, but the overall peak is expected during April 15-20... We can be wrong, but I am reasonably confident that we may not be too wrong," he said.

On a positive note, he said the current steep rise will also be followed by a sharp fall. "We'll cross the peak (during April 15-20) and then there will be very sharp drop over the next 15-20 days," said Agrawal. The national 'super model' had in October last year predicted pandemic can be controlled by early 2021 with "minimal active symptomatic infections" in February. Asked about the possible explanations for the second wave, he said it could be due to gatherings linked to opening of schools, colleges and other working areas where people became comparatively more careless, and the fact that there are some mutations which are spreading faster.

DSC 7215

Patient out of ICU as Italy on the brink

Rome, March 10: Italian doctors celebrated one small victory in their battle against the Covid-19 after a 38-year-old man was moved out of intensive care for the first time since he tested positive February 21 and opened Italy's health care crisis as Patient No. 1.

But in the rest of northern Italy, the virus spread was growing so exponentially that doctors were making comparisons to war-time triage medics deciding who lives, who dies and who gets access to the limited ICU beds. "It's a reasoning that our colleagues make," Dr Guido Giustetto, head of the association of doctors in northern Piedmont, said Monday.

"It becomes dramatic if, rather than doing it under normal situations, they do it because the beds are so scarce that someone might not have access to medical care." The Italian society of anesthesiology and intensive care published 15 ethical recommendations to consider when deciding on ICU admissions during the virus crisis and the ICU shortage. The criteria include the age of the patient and the probability of survival, and not just "first come first served".

The Lombardy government has been scrambling to increase its ICU capacity, converting operating and recovery rooms into isolated wards to treat the 440 critical virus patients currently in need. It has cobbled together 150 more beds in the last two weeks and expects another 150 in the coming week. But it may not be enough. "Unfortunately we're only at the beginning," said Dr Massimo Galli, head of infectious disease at Milan's Sacco hospital. Speaking to Sky-Tg24, Galli noted that the numbers of infections registered in Lombardy last week were similar to those in Wuhan, China in late January. — *AP*

- **POPE FRANCIS** urged Catholic priests on Tuesday to "have the courage" to go out and help those sickened by the novel Covid-19.
- **THE VATICAN'S** Saint Peter's Square and its main basilica were closed to tourists on Tuesday.

"When there are 10 people in the emergency room waiting to get on a ventilator, it is entirely feasible that you would be removing people from ventilators knowing that they are going to die. But you remove people from ventilators when your evaluation of the medical situation suggests that patients are not improving. If a patient is not improving, and it doesn't look like using this scarce resource is a wise investment, then you try it out on another patient who might have better luck," he says.

Deccan Chronicle epaper, Hyderabad edition, March 11, 2020

Age of Coronavirus Deaths

COVID-19 Fatality Rate by AGE:

***Death Rate** = (number of deaths / number of cases) = **probability of dying if infected by the virus** (%). This probability differs depending on the age group. The percentages shown below **do not have to add up to 100%**, as they **do NOT represent share of deaths by age** group. Rather, it represents, for a person in a given age group, the **risk of dying** if infected with COVID-19.

AGE	DEATH RATE confirmed cases	DEATH RATE all cases
80+ years old	**21.9%**	**14.8%**
70-79 years old		**8.0%**
60-69 years old		**3.6%**
50-59 years old		**1.3%**
40-49 years old		**0.4%**
30-39 years old		**0.2%**
20-29 years old		**0.2%**
10-19 years old		**0.2%**
0-9 years old		**no fatalities**

Real-World Decision Making

Who to save by calculating probabilities of survival based on various factors.

https://www.worldometers.info/coronavirus/coronavirus-age-sex-demographics/ and https://www.healthleadersmedia.com/clinical-care/4-ethical-dilemmas-healthcare-organizations-during-covid-19-pandemic

Last accessed: March 20, 2020 and April 09, 2020, respectively.

Demographic characteristics

Among 55,924 laboratory confirmed cases reported as of 20 February 2020, the median age is 51 years (range 2 days-100 years old; IQR 39-63 years old) with the majority of cases (77.8%) aged between 30–69 years. Among reported cases, 51.1% are male, 77.0% are from Hubei and 21.6% are farmers or laborers by occupation.

Below is an extract of the study findings (highlight added by Worldometer):

Among the first 425 patients with confirmed NCIP, the median age was 59 years and 56% were male. The majority of cases (55%) with onset before January 1, 2020, were linked to the Huanan Seafood Wholesale Market, as compared with 8.6% of the subsequent cases.

The mean incubation period was 5.2 days (95% confidence interval [CI], 4.1 to 7.0), with the 95th percentile of the distribution at 12.5 days.

In its early stages, the epidemic doubled in size every 7.4 days. With a mean serial interval of 7.5 days (95% CI, 5.3 to 19), the basic reproductive number was estimated to be 2.2 (95% CI, 1.4 to 3.9).

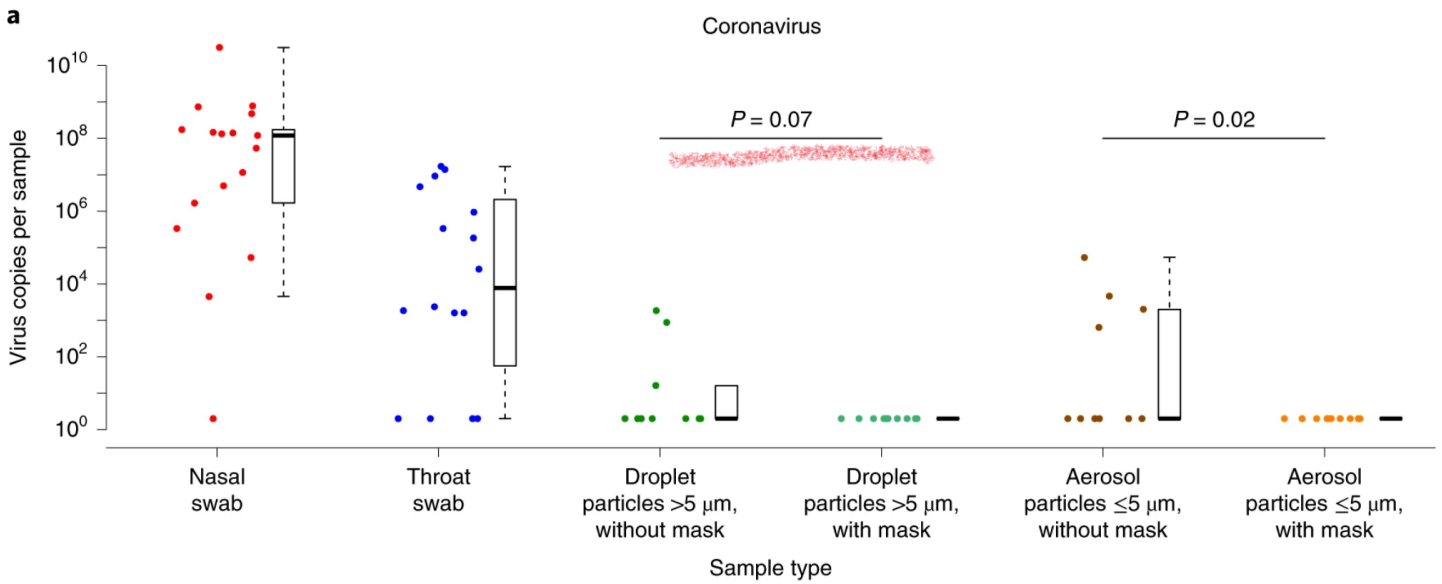
Real-World Decision Making

Calculating descriptive statistics to get an understanding of the data, without which no modeling can be effective.

Source: Report of the WHO-China Joint Report on Coronavirus Disease 2019 (COVID-19), 16-24 February, 2020 and https://www.worldometers.info/coronavirus/coronavirus-incubation-period/#asy
Last accessed: March 20, 2020

Fig. 1: Efficacy of surgical face masks in reducing respiratory virus shedding in respiratory droplets and aerosols of symptomatic individuals with coronavirus, influenza virus or rhinovirus infection.

From: Respiratory virus shedding in exhaled breath and efficacy of face masks



differences in bold. For nasal swabs and throat swabs, all infected individuals were included (coronavirus, n=17; influenza virus, n=43; rhinovirus, n=54). For respiratory droplets and aerosols, numbers of infected individuals who provided exhaled breath samples while not wearing or wearing a surgical face mask, respectively were: coronavirus (n=10 and 11), influenza virus (n=23 and 28) and rhinovirus (n=36 and 32). A subset of participants provided exhaled breath samples for both mask interventions (coronavirus, n=4; influenza virus, n=8; rhinovirus, n=14). The box plots indicate the median with the interquartile range (lower and upper hinge) and ±1.5×interquartile range from the first and third quartile (lower and upper whiskers).

Real-World Decision Making

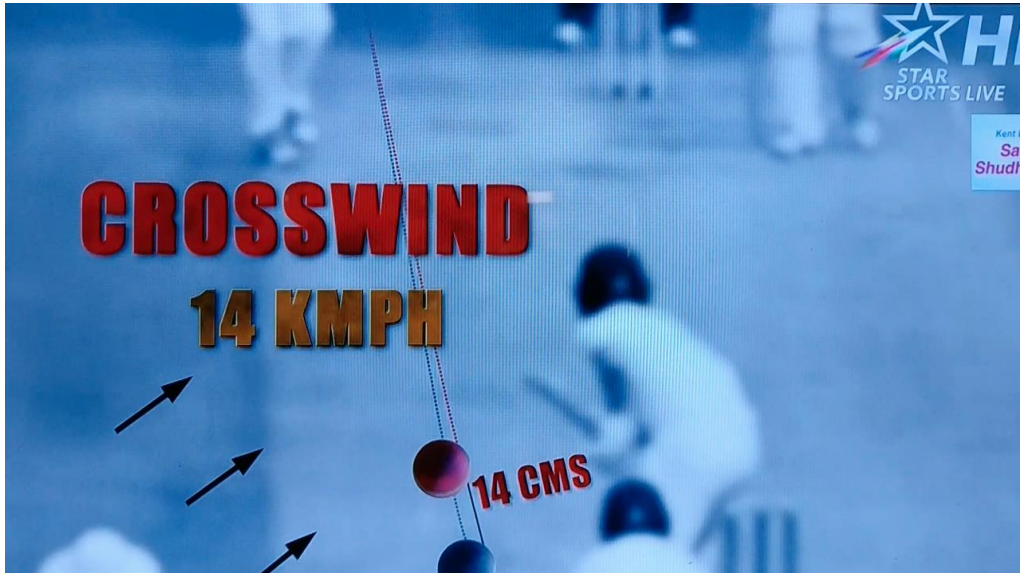
More descriptive statistics and then modeling to determine if face masks are effective in reducing virus shedding.

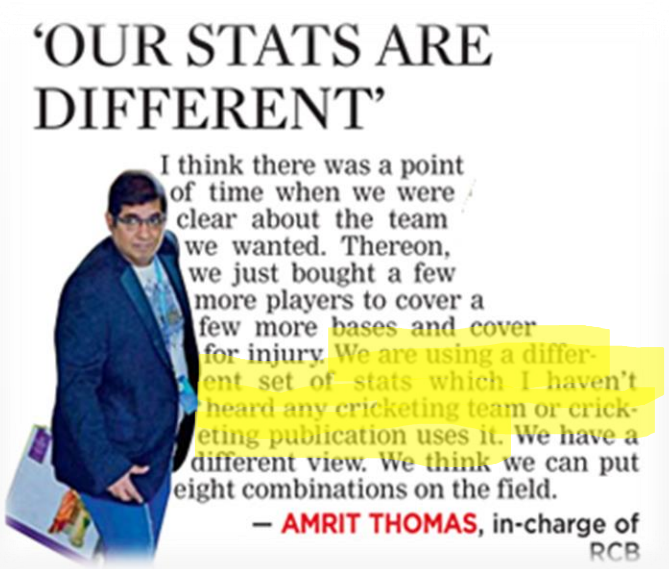
Always Be Aware and Beware!

While decisions should be based on data, you should also understand and question data, and understand assumptions made in building models and/or limitations of data and the models.

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

Statistics in Sports





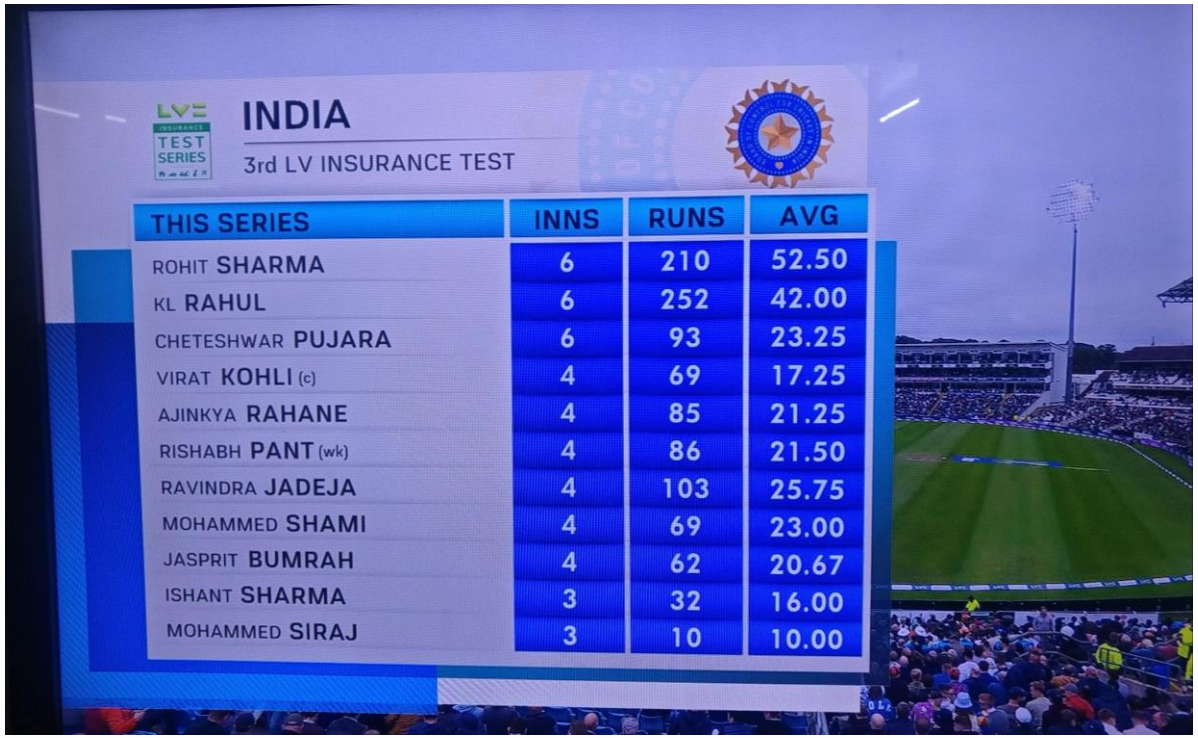
‘OUR STATS ARE DIFFERENT’

I think there was a point of time when we were clear about the team we wanted. Thereon, we just bought a few more players to cover a few more bases and cover for injury. We are using a different set of stats which I haven’t heard any cricketing team or cricketing publication uses it. We have a different view. We think we can put eight combinations on the field.

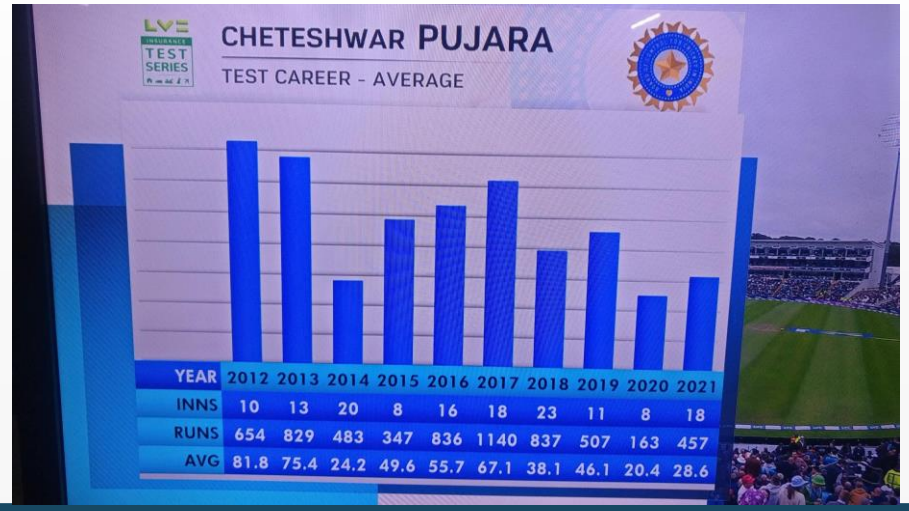
— AMRIT THOMAS, in-charge of RCB

DSC 7215

Statistics in Sports



THIS SERIES	INNS	RUNS	AVG
ROHIT SHARMA	6	210	52.50
KL RAHUL	6	252	42.00
CHETESHWAR PUJARA	6	93	23.25
VIRAT KOHLI (c)	4	69	17.25
AJINKYA RAHANE	4	85	21.25
RISHABH PANT (wk)	4	86	21.50
RAVINDRA JADEJA	4	103	25.75
MOHAMMED SHAMI	4	69	23.00
JASPRIT BUMRAH	4	62	20.67
ISHANT SHARMA	3	32	16.00
MOHAMMED SIRAJ	3	10	10.00

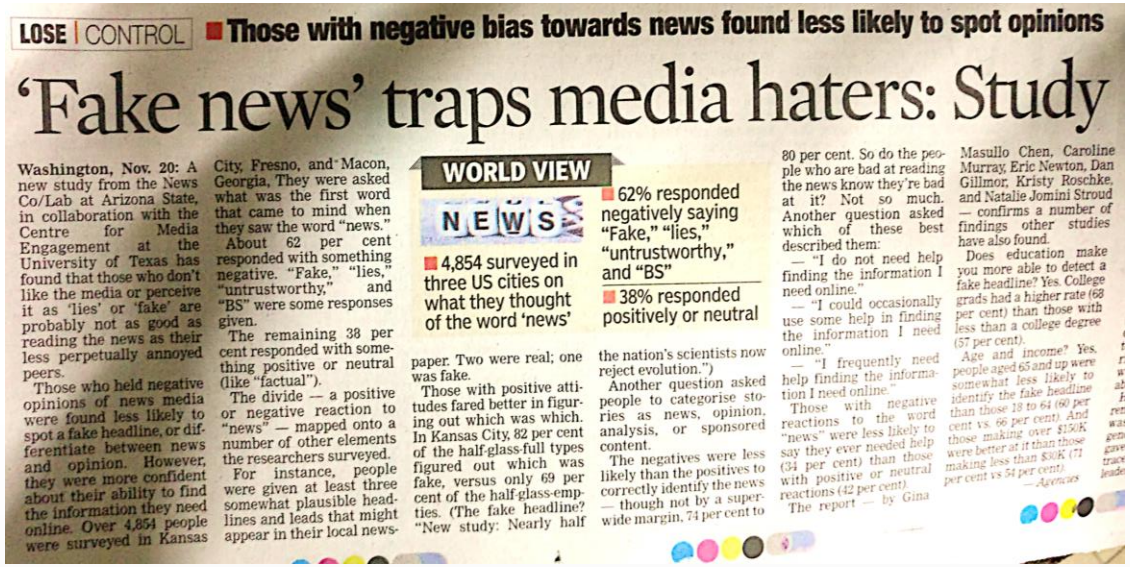


YEAR	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
INNS	10	13	20	8	16	18	23	11	8	18
RUNS	654	829	483	347	836	1140	837	507	163	457
AVG	81.8	75.4	24.2	49.6	55.7	67.1	38.1	46.1	20.4	28.6

Statistics in Politics and Psephology

There are three kinds of lies: lies, damned lies, and statistics.

- Mark Twain / Benjamin Disraeli



LOSE | CONTROL ■ Those with negative bias towards news found less likely to spot opinions

'Fake news' traps media haters: Study

Washington, Nov. 20: A new study from the News Co/Lab at Arizona State, in collaboration with the Centre for Media Engagement at the University of Texas has found that those who don't like the media or perceive it as 'lies' or 'fake' are probably not as good as reading the news as their less perpetually annoyed peers.

Those who held negative opinions of news media were found less likely to spot a fake headline, or differentiate between news and opinion. However, they were more confident about their ability to find the information they need online. Over 4,854 people were surveyed in Kansas City, Fresno, and Macon, Georgia. They were asked what was the first word that came to mind when they saw the word "news."

About 62 per cent responded with something negative. "Fake," "lies," "untrustworthy," and "BS" were some responses given.

The remaining 38 per cent responded with something positive or neutral (like "factual").

The divide — a positive or negative reaction to "news" — mapped onto a number of other elements the researchers surveyed.

For instance, people were given at least three somewhat plausible headlines and leads that might appear in their local newspaper. Two were real; one was fake.

Those with positive attitudes fared better in figuring out which was which. In Kansas City, 82 per cent of the half-glass-full types figured out which was fake, versus only 69 per cent of the half-glass-empties. (The fake headline? "New study: Nearly half the nation's scientists now reject evolution.")

Another question asked people to categorise stories as news, opinion, analysis, or sponsored content.

The negatives were less likely than the positives to correctly identify the news — though not by a super-wide margin. 74 per cent to 80 per cent. So do the people who are bad at reading the news know they're bad at it? Not so much. Another question asked which of these best described them:

— "I do not need help finding the information I need online."

— "I could occasionally use some help in finding the information I need online."

— "I frequently need help finding the information I need online."

Those with negative reactions to the word "news" were less likely to say they ever needed help (34 per cent) than those with positive or neutral reactions (42 per cent).

The report — by Gina Masullo Chen, Caroline Murray, Eric Newton, Dan Gillmor, Kristy Roschke, and Natalie Jomini Stroud — confirms a number of findings other studies have also found.

Does education make you more able to detect a fake headline? Yes. College grads had a higher rate (68 per cent) than those with less than a college degree (57 per cent).

Age and income? Yes, people aged 65 and up were somewhat less likely to identify the fake headline than those 18 to 64 (60 per cent vs 66 per cent). And those making over $150K were better at it than those making less than $30K (71 per cent vs 54 per cent).

— Agencies

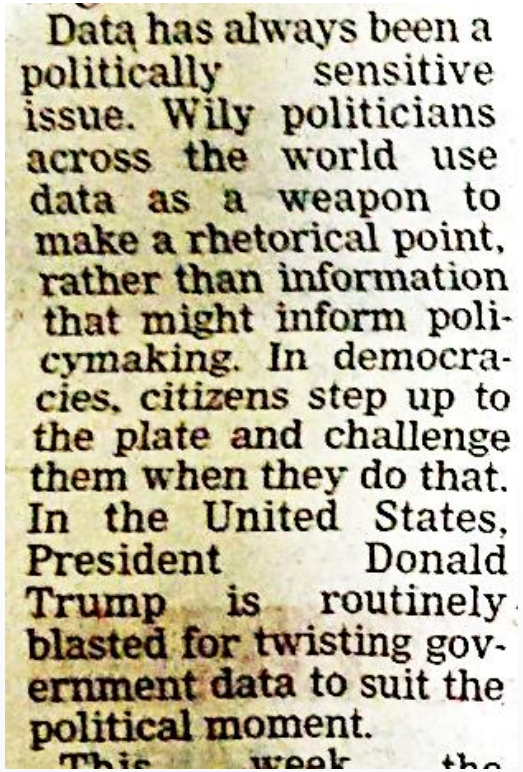
WORLD VIEW

NEWS

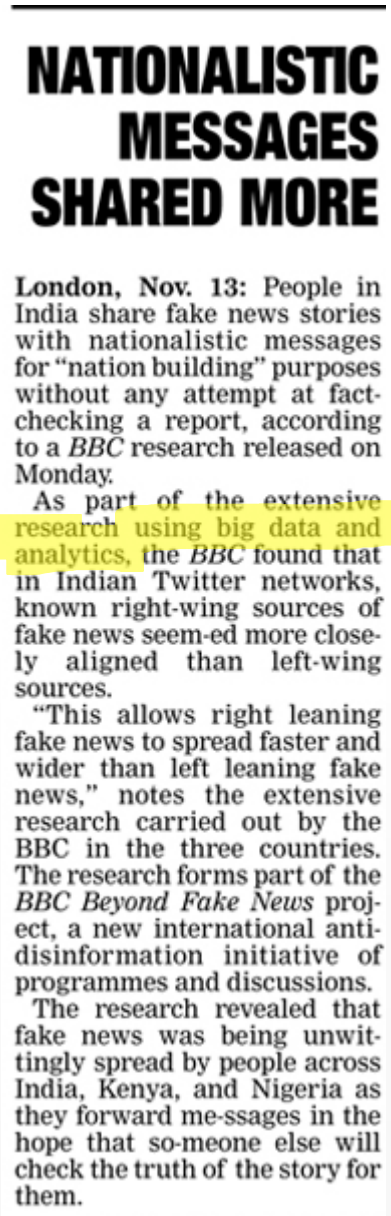
■ 4,854 surveyed in three US cities on what they thought of the word 'news'

■ 62% responded negatively saying "Fake," "lies," "untrustworthy," and "BS"

■ 38% responded positively or neutral



Data has always been a politically sensitive issue. Wily politicians across the world use data as a weapon to make a rhetorical point, rather than information that might inform policymaking. In democracies, citizens step up to the plate and challenge them when they do that. In the United States, President Donald Trump is routinely blasted for twisting government data to suit the political moment.



NATIONALISTIC MESSAGES SHARED MORE

London, Nov. 13: People in India share fake news stories with nationalistic messages for "nation building" purposes without any attempt at fact-checking a report, according to a *BBC* research released on Monday.

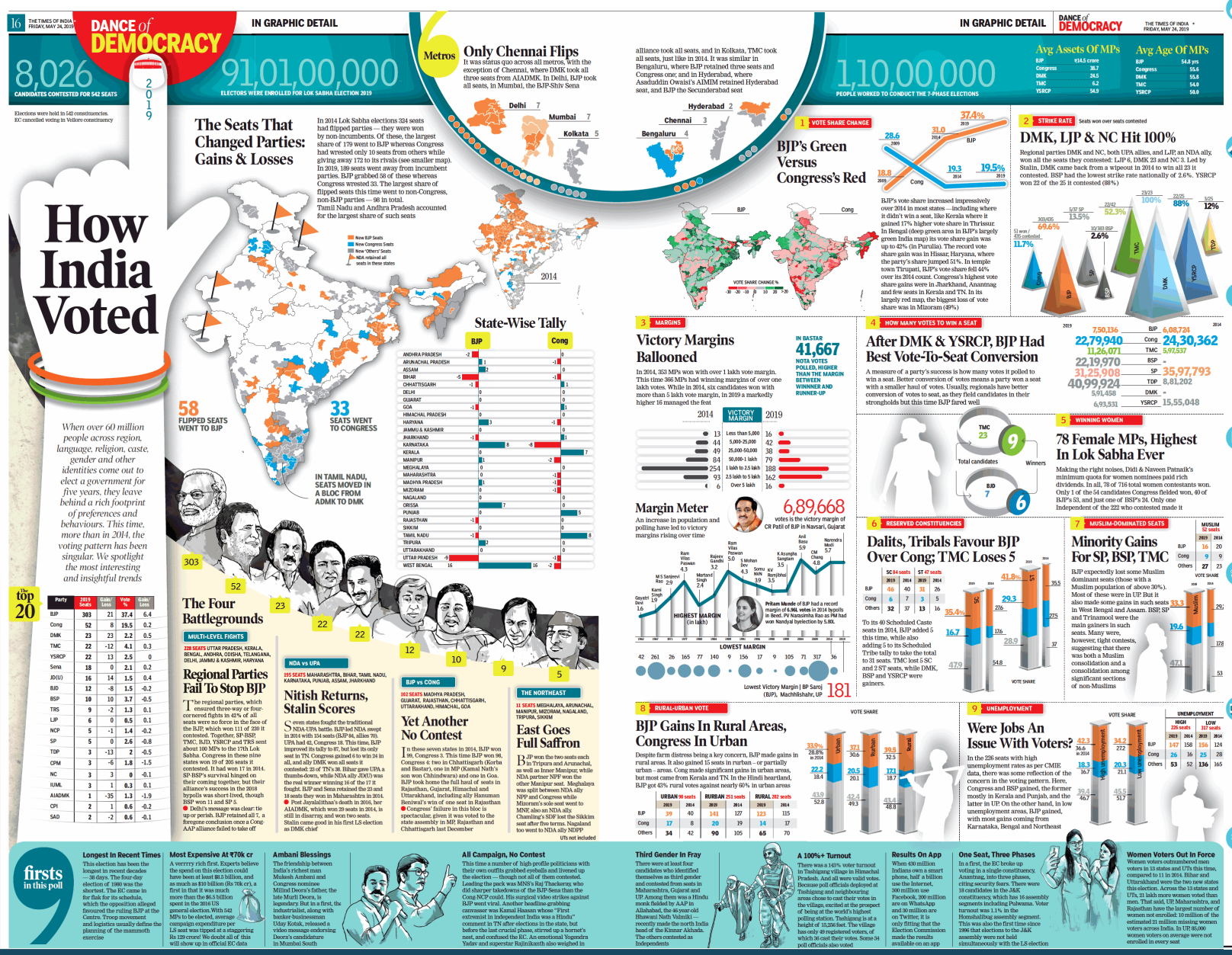
As part of the extensive research using big data and analytics, the *BBC* found that in Indian Twitter networks, known right-wing sources of fake news seem-ed more closely aligned than left-wing sources.

"This allows right leaning fake news to spread faster and wider than left leaning fake news," notes the extensive research carried out by the BBC in the three countries. The research forms part of the *BBC Beyond Fake News* project, a new international anti-disinformation initiative of programmes and discussions.

The research revealed that fake news was being unwittingly spread by people across India, Kenya, and Nigeria as they forward me-ssages in the hope that so-meone else will check the truth of the story for them.

DSC 7215

Statistics in Politics and Psephology


16 THE TIMES OF INDIA FRIDAY, MAY 24, 2019
DANCE of DEMOCRACY
IN GRAPHIC DETAIL
8,026
CANDIDATES CONTESTED FOR 542 SEATS
91,01,00,000
ELECTORS WERE ENROLLED FOR LOK SABHA ELECTION 2019
Elections were held in 542 constituencies. EC cancelled voting in Vellore constituency
2019
How India Voted
When over 60 million people across region, language, religion, caste, gender and other identities come out to elect a government for five years, they leave behind a rich footprint of preferences and behaviours. This time, more than in 2014, the voting pattern has been singular. We spotlight the most interesting and insightful trends
The top 20
The Seats That Changed Parties: Gains & Losses
In 2014 Lok Sabha elections 224 seats had flipped parties — they were won by non-incumbents. Of these, the largest share of 179 went to BJP whereas Congress had wrested only 10 seats from others while giving away 172 to its rivals (see smaller map). In 2019, 189 seats went away from incumbent parties. BJP grabbed 58 of these whereas Congress wrested 33. The largest share of flipped seats this time went to non-Congress, non-BJP parties — 98 in total. Tamil Nadu and Andhra Pradesh accounted for the largest share of such seats
New BJP Seats
New Congress Seats
New 'Others' Seats
NDA retained all seats in these states
2014
58 FLIPPED SEATS WENT TO BJP
33 SEATS WENT TO CONGRESS
IN TAMIL NADU, SEATS MOVED IN A BLOC FROM ADMK TO DMK
State-Wise Tally
BJP
Cong
ANDHRA PRADESH -2 0
ARUNACHAL PRADESH 1 -1
ASSAM 2 0
BIHAR -5 -1
CHHATTISGARH -1 1
DELHI 0 0
GUJARAT 0 0
GOA -1 1
HIMACHAL PRADESH 0 0
HARYANA 3 -1
JAMMU & KASHMIR 0 0
JHARKHAND -1 1
KARNATAKA 8 -8
KERALA 7
MANIPUR 1 -2
MEGHALAYA 0 0
MAHARASHTRA 0 -1
MADHYA PRADESH 1 -1
MIZORAM 0 -1
NAGALAND 0 0
ORISSA 7 0
PUNJAB 0 5
RAJASTHAN -1 0
SIKKIM 0 0
TAMIL NADU -1 8
TRIPURA 2 0
UTTARAKHAND 0 0
UTTAR PRADESH -9 -1
WEST BENGAL 16 16 -2
303
52
23
22
22
12
10
9
5
The Four Battlegrounds
MULTI-LEVEL FIGHTS
228 SEATS UTTAR PRADESH, KERALA, BENGAL, ANDHRA, ODISHA, TELANGANA, DELHI, JAMMU & KASHMIR, HARYANA
Regional Parties Fail To Stop BJP
The regional parties, which ensured three-way or four-cornered fights in 42% of all seats were no force in the face of the BJP, which won 111 of 220 it contested. Together, SP-BSP, TMC, BJD, YSRCP and TRS sent about 100 MPs to the 17th Lok Sabha. Congress in these nine states won 19 of 205 seats it contested. It had won 17 in 2014. SP-BSP's survival hinged on their coming together, but their alliance's success in the 2018 bypolls was short-lived, though BSP won 11 and SP 5.
Delhi's message was clear: tie up or perish. BJP retained all 7, a foregone conclusion once a Cong-AAP alliance failed to take off
NDA vs UPA
195 SEATS MAHARASHTRA, BIHAR, TAMIL NADU, KARNATAKA, PUNJAB, ASSAM, JHARKHAND
Nitish Returns, Stalin Scores
Seven states fought the traditional NDA-UPA battle. BJP-led NDA swept in 2014 with 154 seats (BJP 84, allies 70). UPA had 42, Congress 18. This time, BJP improved its tally to 87, but lost its only seat in TN. Congress gained 6 to win 24 in all, and ally DMK won all seats it contested: 23 of TN's 38. Bihar gave UPA a thumbs-down, while NDA ally JD(U) was the real winner winning 16 of the 17 it fought. BJP and Sena retained the 23 and 18 seats they won in Maharashtra in 2014.
Post Jayalalithaa's death in 2016, her AIADMK, which won 29 seats in 2014, is still in disarray, and won two seats. Stalin came good in his first LS election as DMK chief
BJP vs CONG
102 SEATS MADHYA PRADESH, GUJARAT, RAJASTHAN, CHHATTISGARH, UTTARAKHAND, HIMACHAL, GOA
Yet Another No Contest
In these seven states in 2014, BJP won 99, Congress 4: two in Chhattisgarh (Korba and Raatar), one in MP (Kamal Nath's son won Chhindwara) and one in Goa. BJP took home the full haul of seats in Rajasthan, Gujarat, Himachal and Uttarakhand, including ally Hanuman Beniwal's win of one seat in Rajasthan
Congress' failure in this bloc is spectacular, given it was voted to the state assembly in MP, Rajasthan and Chhattisgarh last December
THE NORTHEAST
11 SEATS MEGHALAYA, ARUNACHAL, MANIPUR, MIZORAM, NAGALAND, TRIPURA, SIKKIM
East Goes Full Saffron
BJP won the two seats each in Tripura and Arunachal, as well as Inner Manipur, while NDA partner NPF won the other Manipur seat. Meghalaya was split between NDA ally NPP and Congress while Mizoram's sole seat went to MNF, also an NDA ally. Chamling's SDF lost the Sikkim seat after five terms. Nagaland too went to NDA ally NDPP
firsts in this poll
Longest In Recent Times
This election has been the longest in recent decades — 38 days. The four-day election of 1980 was the shortest. The EC came in for flak for its schedule, which the opposition alleged favoured the ruling BJP at the Centre. Troop movement and logistics usually define the planning of the mammoth exercise
Most Expensive At ₹70k cr
A verrrry rich first. Experts believe the spend on this election could have been at least $6.5 billion, and as much as $10 billion (Rs 70k cr), a first in that it was much more than the $6.5 billion spent in the 2016 US general election. With 542 MPs to be elected, average campaign expenditure per LS seat was tipped at a staggering Rs 129 crore! We doubt all of this will show up in official EC data
Ambani Blessings
The friendship between India's richest man Mukesh Ambani and Congress nominee Milind Deora's father, the late Murli Deora, is legendary. But in a first, the industrialist, along with banker-businessman Uday Kotak, released a video message endorsing Deora's candidature in Mumbai South
All Campaign, No Contest
This time a number of high-profile politicians with their own outfits grabbed eyeballs and livened up the election — though not all of them contested. Leading the pack was MNS's Raj Thackeray, who did sharper takedowns of the BJP-Sena than the Cong-NCP could. His surgical video strikes against BJP went viral. Another headline-grabbing cameoer was Kamal Haasan whose "First extremist in independent India was a Hindu" comment in TN after elections in the state, but before the last crucial phase, stirred up a hornet's nest, and confused the EC. An emotional Yogendra Yadav and superstar Rajinikanth also weighed in
Metros
Only Chennai Flips
It was status quo across all metros, with the exception of Chennai, where DMK took all three seats from AIADMK. In Delhi, BJP took all seats, in Mumbai, the BJP-Shiv Sena alliance took all seats, and in Kolkata, TMC took all seats, just like in 2014. It was similar in Bengaluru, where BJP retained three seats and Congress one; and in Hyderabad, where Asaduddin Owaisi's AIMIM retained Hyderabad seat, and BJP the Secunderabad seat
Delhi 7
Mumbai 7
Kolkata 5
Hyderabad 2
Chennai 3
Bengaluru 4
IN GRAPHIC DETAIL
DANCE of DEMOCRACY
THE TIMES OF INDIA FRIDAY, MAY 24, 2019
1,10,00,000
PEOPLE WORKED TO CONDUCT THE 7-PHASE ELECTIONS
Avg Assets Of MPs
BJP ₹14.5 crore
Congress 38.7
DMK 24.5
TMC 6.2
YSRCP 54.9
Avg Age Of MPs
BJP 54.8 yrs
Congress 55.6
DMK 55.8
TMC 54.0
YSRCP 58.0
1 VOTE SHARE CHANGE
BJP's Green Versus Congress's Red
28.6 2009
31.0 2014
37.4% 2019 BJP
18.8 2009
19.3 2014
19.5% 2019 Cong
BJP
Cong
VOTE SHARE CHANGE %
BJP's vote share increased impressively over 2014 in most states — including where it didn't win a seat, like Kerala where it gained 17% higher vote share in Thrissur. In Bengal (deep green area in BJP's largely green India map) its vote share gain was up to 42% (in Purulia). The record vote share gain was in Hissar, Haryana, where the party's share jumped 51%. In temple town Tirupati, BJP's vote share fell 44% over its 2014 count. Congress's highest vote share gains were in Jharkhand, Anantnag and few seats in Kerala and TN. In its largely red map, the biggest loss of vote share was in Mizoram (49%)
2 STRIKE RATE
Seats won over seats contested
DMK, LJP & NC Hit 100%
Regional parties DMK and NC, both UPA allies, and LJP, an NDA ally, won all the seats they contested: LJP 6, DMK 23 and NC 3. Led by Stalin, DMK came back from a wipeout in 2014 to win all 23 it contested. BSP had the lowest strike rate nationally of 2.6%. YSRCP won 22 of the 25 it contested (88%)
51 won / 435 contested 11.7% Cong
30/435 69.6% BJP
5/72 SP 13.5%
10/383 BSP 2.6%
22/42 52.3% SP
BSP
29/29 100% TMC
22/25 88% DMK
2/25 12% YSRCP
TDP
3 MARGINS
Victory Margins Ballooned
In 2014, 353 MPs won with over 1 lakh vote margin. This time 366 MPs had winning margins of over one lakh votes. While in 2014, six candidates won with more than 5 lakh vote margin, in 2019 a markedly higher 16 managed the feat
IN BASTAR 41,667 NOTA VOTES POLLED, HIGHER THAN THE MARGIN BETWEEN WINNER AND RUNNER-UP
VICTORY MARGIN
2014 / 2019
13 / 16 Less than 5,000
44 / 42 5,000-25,000
49 / 38 25,000-50,000
84 / 79 50,000-1 lakh
254 / 188 1 lakh to 2.5 lakh
93 / 162 2.5 lakh to 5 lakh
6 / 16 Over 5 lakh
Margin Meter
An increase in population and polling have led to victory margins rising over time
6,89,668 votes is the victory margin of CR Patil of BJP in Navsari, Gujarat
Gayatri Devi 1.6
Kanti Singh 1.9
M S Sanjeevi Rao 2.9
Ram Vilas Paswan 4.3
Mahtand Singh 2.4
Rajeev Gandhi 3.2
Ram Vilas Paswan 5.0
S Mohan Dev 4.3
Somu NVR 3.9
K Arunghia Sangkam 3.5
Anil Basu 5.9
OM Chang 4.8
Narendra Modi 5.7
HIGHEST MARGIN (in lakh)
Priram Munde of BJP had a record margin of 6.96L votes in 2014 bypolls in Beed. PV Narasimha Rao so PM had won Nandyal byelection by 5.8L
LOWEST MARGIN
42 261 26 165 77 140 9 156 17 9 105 71 317 36
Lowest Victory Margin | BP Saroj (BJP), Machhlishahr, UP 181
4 HOW MANY VOTES TO WIN A SEAT
After DMK & YSRCP, BJP Had Best Vote-To-Seat Conversion
A measure of a party's success is how many votes it polled to win a seat. Better conversion of votes means a party won a seat with a smaller haul of votes. Usually, regionals have better conversion of votes to seat, as they field candidates in their strongholds but this time BJP fared well
2019 / 2014
7,50,136 BJP 6,08,724
22,79,940 Cong 24,30,362
11,26,071 TMC 5,97,537
22,19,970 BSP -
31,25,908 SP 35,97,793
40,99,924 TDP 8,81,202
5,91,458 DMK -
6,93,531 YSRCP 15,55,048
TMC 23 9 Total candidates Winners
BJD 7 6
5 WINNING WOMEN
78 Female MPs, Highest In Lok Sabha Ever
Making the right noises, Didi & Naveen Patnaik's minimum quota for women nominees paid rich dividends. In all, 78 of 716 total women contestants won. Only 1 of the 54 candidates Congress fielded won, 40 of BJP's 53, and just one of BSP's 24. Only one Independent of the 222 who contested made it
6 RESERVED CONSTITUENCIES
Dalits, Tribals Favour BJP Over Cong; TMC Loses 5
SC 84 seats / ST 47 seats
2019 / 2014 / 2019 / 2014
BJP 46 40 31 26
Cong 6 7 3 5
Others 32 37 13 16
To its 40 Scheduled Caste seats in 2014, BJP added 5 this time, while also adding 5 to its Scheduled Tribe tally to take the total to 31 seats. TMC lost 5 SC and 2 ST seats, while DMK, BSP and YSRCP were gainers.
SC 35.4% 16.7 47.9
ST 41.8% 29.3 28.9 54.8
VOTE SHARE
7 MUSLIM-DOMINATED SEATS
Minority Gains For SP, BSP, TMC
BJP expectedly lost some Muslim dominant seats (those with a Muslim population of above 30%). Most of these were in UP. But it also made some gains in such seats in West Bengal and Assam. BSP, SP and Trinamool were the main gainers in such seats. Many were, however, tight contests, suggesting that there was both a Muslim consolidation and a consolidation among significant sections of non-Muslims
MUSLIM 52 seats
2019 / 2014
BJP 16 20
Cong 9 9
Others 27 23
VOTE SHARE
Muslim 33.3 19.6 47.1
Non-Muslim 29.2 17.8 53
8 RURAL-URBAN VOTE
BJP Gains In Rural Areas, Congress In Urban
Despite farm distress being a key concern, BJP made gains in rural areas. It also gained 15 seats in rurban — or partially urban — areas. Cong made significant gains in urban areas, but most came from Kerala and TN. In the Hindi heartland, BJP got 43% rural votes against nearly 60% in urban areas
URBAN 90 seats / RURBAN 251 seats / RURAL 202 seats
2019 / 2014 / 2019 / 2014 / 2019 / 2014
BJP 39 40 141 127 123 115
Cong 17 8 20 19 14 17
Others 34 42 90 105 65 70
UTs not included
VOTE SHARE
Urban 33.9% 28.8% in 2018 22.2 18.4 43.9 52.8
Rurban 37.1 30.6 20.5 20.1 42.4 49.3
Rural 39.5 32.5 17.1 18.7 43.4 48.8
9 UNEMPLOYMENT
Were Jobs An Issue With Voters?
In the 226 seats with high unemployment rates as per CMIE data, there was some reflection of the concern in the voting pattern. Here, Congress and BSP gained, the former mostly in Kerala and Punjab, and the latter in UP. On the other hand, in low unemployment areas, BJP gained, with most gains coming from Karnataka, Bengal and Northeast
VOTE SHARE
High unemployment 42.3 36.6 in 2014 18.3 16.7 39.4 46.7
Low unemployment 34.2 27.2 20.3 21.1 45.5 51.7
UNEMPLOYMENT
HIGH 226 seats / LOW 317 seats
2019 / 2014 / 2019 / 2014
BJP 147 158 156 124
Cong 26 16 25 28
Others 53 52 136 165
Women Voters Out In Force
Women voters outnumbered men voters in 13 states and UTs this time, compared to 11 in 2014. Bihar and Uttarakhand were the two new states this election. Across the 13 states and UTs, 21 lakh more women voted than men. That said, UP, Maharashtra, and Rajasthan have the largest number of women not enrolled: 10 million of the estimated 21 million missing women voters across India. In UP, 18,000 women voters on average were not enrolled in every seat
Third Gender In Fray
There were at least four candidates who identified themselves as third gender and contested from seats in Maharashtra, Gujarat and UP. Among them was a Hindu monk fielded by AAP in Allahabad, the 49-year-old Bhawani Nath Valmiki — recently made the north India head of the Kinnar Akhada. The others contested as Independents
A 100%+ Turnout
There was a 143% voter turnout in Tashigang village in Himachal Pradesh. And all were valid votes. Because poll officials deployed at Tashigang and neighbouring areas chose to cast their votes in the village, excited at the prospect of being at the world's highest polling station. Tashigang is at a height of 15,256 feet. The village has only 49 registered voters, of which 36 cast their votes. Some 34 poll officials also voted
Results On App
When 430 million Indians own a smart phone, half a billion use the Internet, 300 million use Facebook, 200 million are on WhatsApp and 30 million are on Twitter, it is only fitting that the Election Commission made the results available on an app
One Seat, Three Phases
In a first, the EC broke up voting in a single constituency, Anantnag, into three phases, citing security fears. There were 18 candidates in the J&K constituency, which has 18 assembly segments including Pulwama. Voter turnout was 1.1% in the Homshalbnag assembly segment. This was also the first time since 1996 that elections to the J&K assembly were not held simultaneously with the LS election

Modi & Key Words

An analysis of words reflecting key issues or politics in the PM's speeches between July 2014 and April 2019

Chowkidar: A word that flooded politics after Rahul Gandhi's jibe. Modi looked to turn the tables by prefixing 'chowkidar' to his Twitter handle, positioning himself as the country's watchman before elections. The word appeared more frequently in Modi's speeches from January. But it still wasn't much, around 6 mentions every 10,000

Mentions per 100 words
Highest frequency: Apr 2019
Roughly 1 mention every 80 sentences
Jul'14 Oct'15 Jan'17 Apr'19

Vikas: Through his tenure, this was one of the PM's favourite words, emphasising his govt's development agenda. The number of mentions every 10,000 words remained between 20 and 28 almost consistently. But it dipped as the election campaign began, falling to around 19 in April 2019 (roughly once every 26 sentences)

Jul'14 Jan'16 Jan'17 Apr'18 Apr'19

Farmers: Mentions of 'khet' (farm land), 'kisan' (farmer), 'krishi' (agriculture), 'fasal' (crop) and 'khad' (fertilizer) increased as farm distress became more talked about. In April 2016 and April 2017, the usage peaked, to over 40 times every 10,000 words. In April 2019, during the campaign, it fell to around 26 times (a mention every 19 sentences)

Jul'14 Apr'15 Apr'16 Apr'17 Apr'18 Apr'19

Congress: You hear it very often in Modi speeches during elections. But he hardly mentions Congress otherwise. In Oct 2016, for instance, he was mentioning Congress thrice every 1L words. It rose to a peak of around 28 mentions every 10,000 words in Jan this year (a mention every 17 sentences)

Jul'14 Oct'14 Jan'17 Apr'19

GST: Mentions of GST touched a high of around 9 every 10,000 words in October 2017, just after its launch in July. But by April 2019, when the election campaign had begun, GST was out in the cold: the PM was only mentioning it around 6 times every 1L words (a mention every 830 sentences)

Jul'14 Oct'16 Oct'17 Apr'19

Demonetisation: Words like 'black money', 'note' and 'kala dhan' entered the PM's speeches from October 2016. That was just a month before demonetisation. It peaked to around 15 mentions every 10,000 words in Jan 2017 (a mention every 33 sentences) but tailed off soon and was used sparingly thereafter

Jul'14 Jan'17 Jan'18 Apr'19

Jobs: Mention of 'jobs', 'employment', 'rozgar' and 'naukri' in the PM's speeches remained consistent in the first three years of his tenure (12 mentions per 10,000 words in April 2016 and '17) but fell after that. During the Lok Sabha campaign, mentions fell to around 5 times per 10,000 words (once every 100)

Jul'14 Apr'16 Apr'17 Apr'18 Apr'19

Calculation based on the estimate that 20 words make a sentence

DSC 7215

Source: Times of India, Mumbai edition, May 24, 2019
Last accessed: May 30, 2019

…The girls, we say, do not have to earn; so why should they be educated?…

Statistics in Economy

INDIA BEATS CHINA, GROWS 7.5% IN Q4

Economists doubtful; India Inc sees downside risks

PAWAN BALI | DC
NEW DELHI, MAY 29

India's GDP accelerated in January to March 2015 period (fourth quarter of 2014-15) to 7.5 per cent, overtaking China as the fastest growing major economy in the world.

This took the overall GDP growth for FY 2014-15 to 7.3 per cent against 6.9 per cent in the 2013-14.

Finance minister Arun Jaitley said that the economy is clearly on "a recovery path". He said an economy growing at fastest pace in the world cannot be 'fragile' as alleged by the former Prime Minister Manmohan Singh.

Finance ministry said that those sectors within control of policy — manufacturing and services — improved substantially while those dependent on factors beyond the policy control such as agriculture (dependent on weather) and exports (on foreign demand), "did less well."

However, the high GDP numbers have come on the back of new methodology which the Central Statistical Organisation (CSO) adopted earlier this year to calculate the GDP.

In January, CSO revised India's growth to 6.9 per cent in the 2013-14 from the earlier estimate of 4.7 per cent.

Before this data revision, it was a dominant view that India had seen worst growth in 2013-14.

Some economists said that the latest GDP numbers released on Friday does not reflect the ground reality and people will now have to adjust that 7 per cent plus growth under the new methodology is not the same what it was based on older formula.

While the GDP has grown at a fast pace the corporate earnings are dismal, industrial activity is weak, private investment is yet to start and bank credit uptake is slow. Rating agency, India ratings, principal economist and director (public finance) Sunil Kumar Sinha said that the people need to recognise that methodology to calculate GDP has undergone a change and it is no longer estimated at factor cost but at market cost.

HIGHLIGHTS

- INDIA'S GDP ACCELERATED TO 7.5 PER CENT, OVERTAKING CHINA AS THE FASTEST GROWING MAJOR ECONOMY IN THE WORLD.
- FACTORS BEYOND THE POLICY CONTROL SUCH AS AGRICULTURE (DEPENDENT ON WEATHER) AND EXPORTS (ON FOREIGN DEMAND), "DID LESS WELL."
- AGRICULTURE GROWTH DECLINED FROM 3.7 PER CENT IN 2013-14 TO 0.2 PER CENT IN 2014-15 DUE TO A WEAK MONSOON. SERVICES GROWTH INCREASED TO 10.2 PER CENT IN 2014-15.

We are not cheerleader,

Mumbai, June 2: Doing some plain-speaking, governor Raghuram Rajan on Tuesday said RBI is not a 'cheerleader' and he chose to "err a bit" in lowering the rates to push investments as growth on ground may not be as high as what is suggested by the headline GDP numbers.

Wondering why an economy needed rate cut when it was growing at 7.5 per cent, Dr Rajan said there is a "contradiction" between the high GDP numbers and the poor corporate earnings along with lack of any visible pick-up in demand.

"In some sense it is a Goldilocks policy, just right given the current situation," he said, as he defended his third 0.25 per cent rate cut this year despite lingering concerns over the below-normal monsoons as well as steadily firming oil prices and their impact on inflation. — *PTI*

India's GDP growth slows, loses world's fastest number 1 tag

MADHUSUDAN SAHOO
NEW DELHI, MAY 31

With a slowdown seen in the key sectors like agriculture, industry and manufacturing in the past nine months that triggered unemployment growth rate to some extent, India's economic growth rate slipped to five-year low of 5.8 per cent in January-March 2018-19.

Amid lower public spending and consumption demand, the government confirmed the sense of slowdown fears among *aam admi* on the eve of assuming its office with some expectations of a wide-ranging policy impetus to turnaround the country's economy. The data came on the day when Nirmala Sitharaman took charge as the new finance minister in the second innings of the Narendra Modi government.

According to data released by the Central Statistics Office (CSO) on Friday, India's gross domestic product or GDP growth slowed to 5.8 per cent in the January-March quarter of FY2018-19, the lowest in 17 quarters, as against 6.6 per cent in the December quarter. With this data India lost fastest-growing large economy tag to China for the first time in two-year as China's economy posted 6.4 per cent growth in the March quarter.

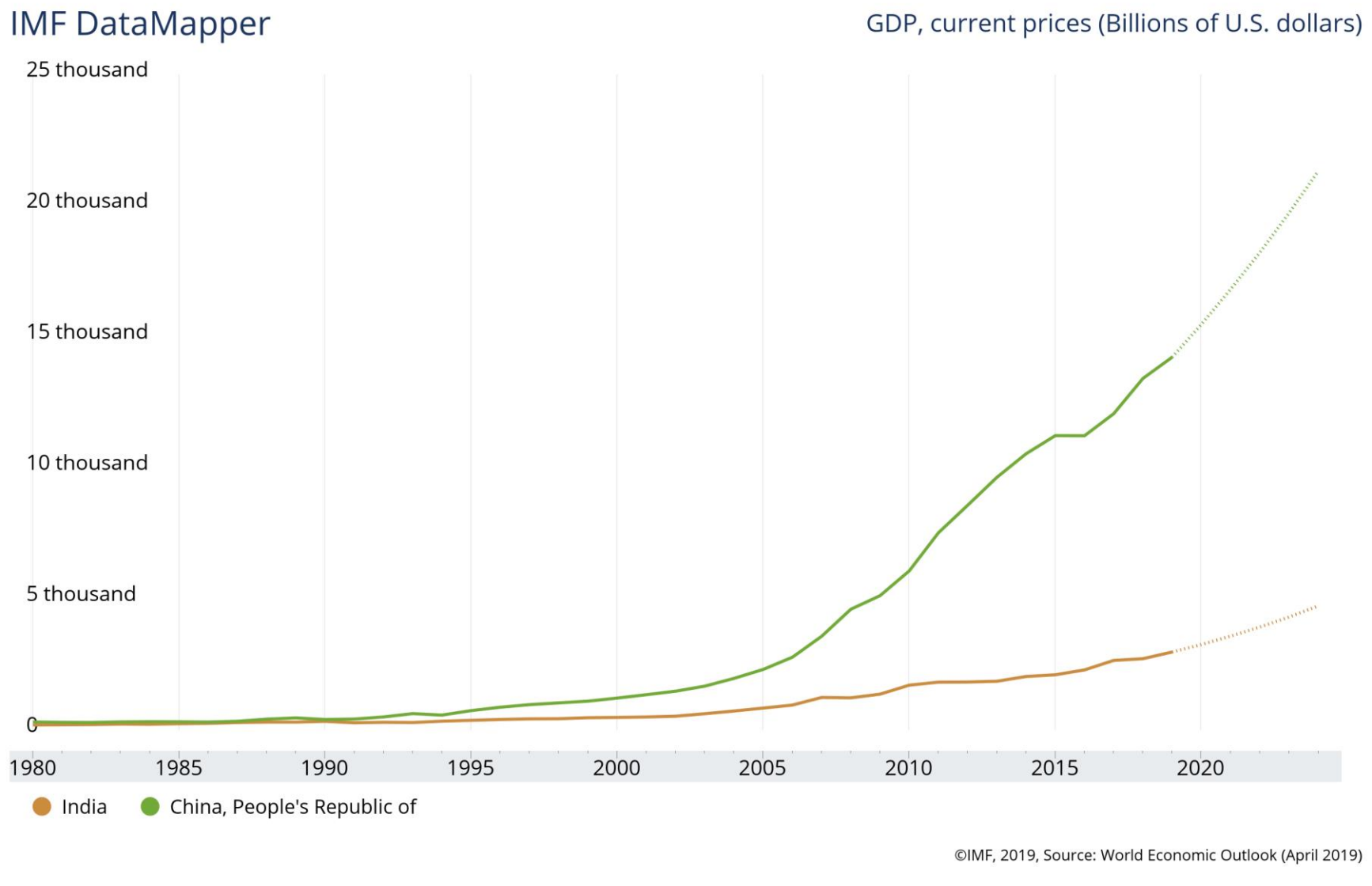
Reacting the slowdown in Q4 GDP, economic affairs secretary S.C. Garg said that it was due to temporary factors like stress in NBFC sector. Mr Garg is also hopeful that economy will be on an upward movement from second quarter onward. "I think first quarter of the current fiscal may also witness relatively slow growth, but economy to pick up second quarter onward," said Garg.

The CSO also revealed that GDP growth during 2018-19 fiscal stood at 6.8 per cent, lower than 7.2 per cent in the previous financial year.

"The growth in GDP was slowest since 2014-15. The previous low was 6.4 per cent in 2013-14. The fourth quarter growth was below China's 6.4 per cent," the data said.

For more on the topic: https://www.quora.com/How-is-the-new-method-to-calculate-GDP-different-from-how-it-was-calculated-earlier-Will-the-other-countries-agree-with-our-new-data-Does-it-mean-that-every-country-calculates-GDP-differently
Last accessed: April 12, 2018

Statistics in Economy



Source: https://www.imf.org/external/datamapper/NGDPD@WEO/IND/CHN?year=2018
Last accessed: June 03, 2019

RAJAN: INDIA SENSITIVE TO GROWTH

DC CORRESPONDENT
WASHINGTON, NOV. 10

Demonetisation and the Goods and Services Tax (GST) are the two major headwinds that held back India's economic growth last year, former RBI Governor Raghuram Rajan has said, asserting that the current seven per cent growth rate is not enough to meet the country's needs.

Mr Rajan, in his address, said a growth rate of seven per cent per year for 25 years is "very very strong" growth, but in some sense this has become the new Hindu rate of growth, which earlier used to be three-and-a-half per cent, Rajan said.

"The reality is that seven is not enough for the kind of people coming into the labour market and we need jobs for them. So, we need more and cannot be satisfied at this level," he said. Observing that India is sensitive to global growth, he said India has become a much more open economy, and if the world grows, it also grows more. "What happened in 2017 is that even as the world picked up, India went down. That reflects

Statistics in Healthcare

GOOD NEWS

Cholesterol not a threat

US to remove high-cholesterol food from 'naughty' list

Washington, May 26: United States officials have finally given the green light for a U-turn on previous warnings on cholesterol, which has been on the "naughty" list of nutrients for nearly 40 years. Health officials have been warning people to stay away from high-cholesterol foods since the 1970s to avoid heart disease and clogged arteries.

However, after a study, eggs, butter, full-fat dairy products, nuts, coconut oil and meat have now been classified as "safe" and have been officially removed from the "nutrients of concern" list, reported the *Internation Business Times*.

The US Department of Agriculture, which is responsible for updating the guidelines every five years, stated in its findings for 2015: "Previously, the Dietary Guidelines for Americans recommended that cholesterol intake be limited to no more than 300 mg/day. The 2015 DGAC will not bring forward this recommendation because available evidence shows no appreciable relationship between consumption of dietary cholesterol and serum (blood) cholesterol, consistent with the AHA/ACC (American Heart Association / American College of Cardiology) report. Cholesterol is not a nutrient of concern for overconsumption."

The Dietary Guidelines Advisory Committee will, in response, no longer warn people against eating high-cholesterol foods and will instead focus on sugar as the main substance of dietary concern.

The 70s, 80s and 90s were the 'non fat' years, with the US government warning people to limit the amount of high-cholersterol foods in their diets to avoid heart disease and strokes.

But nutritionists and scientists have long been campaigning for the U-turn, which started with introducing "good cholesterols" back into the 'safe zone'.

US cardiologist Dr Steven Nissen said: "It's the right decision. We got the dietary guidelines wrong. They've been wrong for decades."

Dr Chris Masterjohn added: "When we eat more foods rich in this compound, our bodies make less. If we deprive ourselves of foods high in cholesterol — such as eggs, and butter — our body revs up its cholesterol synthesis." — *Agencies*

FOODIES' DELIGHT

Butter, full-fat dairy products, nuts, coconut oil and meat have now been classified as "safe" and have been officially removed from the "nutrients of concern" list.

The 70s, 80s and 90s were the 'non fat' years, with the US government warning people to limit the amount of high-cholersterol foods in their diets.

False negative reports now pose a big challenge

Sumitra Deb Roy & Malathy Iyer

Mumbai: A doctor who developed a dry cough and fever was relieved to test negative for Covid-19 at the BMC-run Kasturba Hospital recently. However a couple of days later, he tested positive at a private hospital he drove himself to, due to worsening of his symptoms.

False negative reports are clearly emerging as an add

UP TO 30% SUCH RESULTS IN WUHAN

- A false negative result means a patient's test result is "negative" even though he harbours the virus
- Such patients get sent back from hospitals, increasing the chances of spreading the infection
- Studies from Wuhan, where the Covid-19 pandemic began, found false negative reports could be as high as 30%
- Experts cite a varying reasons—the nature of the virus, the incubation period of the infection, improper collection of samples, types of kits available etc

SECOND TEST CARRIED OUT TO ENSURE NO FALSE POSITIVES AT NIV

From Page 1

The second test is carried out at NIV to ensure that there are no false positives.

WHO looking into reports of people testing -ve, then +ve

Feluda, the first desi Covid detector, gets regulator nod

Will Diagnose Virus In 2 Hrs, To Be Low-Cost

Vishwa Mohan & Rupali Mukherjee | TNN

New Delhi/Mumbai: An indigenous test that uses the cutting-edge CRISPR gene-editing technology to detect the SARS-CoV-2 virus accurately, cheaply and quickly has got the nod for commercial launch from the Drugs Controller General of India.

Developed by scientists at the Council of Scientific and Industrial Research's IGIB (Institute of Genomics and Integrative Biology) in Delhi, along with the Tata Group, the test (named Feluda) can detect the virus in two hours.

"The Tata CRISPR test is the world's first diagnostic test to deploy a specially adapted Cas9 protein to successfully detect the virus causing Covid-19," said a government statement.

When the pandemic began, the IGIB team of Debojyoti Chakraborty and Souvik Maiti was working on developing a genome-based diagnostic tool for sickle cell.

They took less than 100 days to "harness" their work to develop Feluda, said IGIB director Dr Anurag Agrawal.

A press release issued by the ministry of science and technology said the test has 96% sensitivity and 98% specificity for detecting the novel coronavirus. CSIR officials told **TOI** that it is similar to a pregnancy test, containing a strip of paper that changes colour on detection of the virus in the sample.

"This marks a significant achievement for the Indian scientific community, moving from R&D to a high-accuracy, scalable and reliable test in less than 100 days," said the statement.

▶Continued on P9

> This marks a significant achievement for the Indian scientific community, moving from R&D to a high-accuracy, scalable and reliable test in less than 100 days
> — Ministry of Science & Tech in a statement

ping a genome-based diagnostic tool for sickle cell.

They took less than 100 days to "harness" their work to develop Feluda, said IGIB director Dr Anurag Agrawal.

A press release issued by the ministry of science and technology said the test has 96% sensitivity and 98% specificity for detecting the novel coronavirus. CSIR officials told **TOI** that it is similar to a pregnancy test, containing a strip of paper that changes colour on detection of the virus in the sample.

"This marks a significant achievement for the In-

Data says India cannot lower guard against Covid-19 yet

▶From P 1

On September 17, the seven-day average of daily cases had reached 93,199 — the highest since the outbreak of the disease. Since then it has fallen to 85,131, according to data available with

Comparison with other countries shows that cases have risen steadily after long stretches of continuous falls. For example, the US reached its first peak on April 11

of cases will definitely increase in India. Among the 10 countries with the highest number of total cases, Brazil, Colombia, Mexico and South Africa haven't so far seen a second peak once their daily averages started falling. Brazil and South Africa reached

DSC 7215

Statistics in Meteorology

Forecast predicts 17% probability of drought in '19

▶Continued from page 1

The forecast suggests 2019's southwest monsoon rainfall is likely to be near normal. Quantitatively, the monsoon seasonal (June to September) rainfall is likely to be 96% of the long period average (LPA) with a model error of ±5%," said M Rajeevan, ministry of earth sciences (MoES) secretary.

IMD chief K J Ramesh explained that both term (normal and near normal) were technically the same. The forecast percentage of 96% sits on the line separating normal from below normal.

The LPA of the seasonal rainfall over the country as a whole for the period 1951-2000 is 89 cm. The monsoon is categorised as "normal" when the June-September period has rainfall in the range of 96-104% of the LPA. On the other hand, the seasonal rainfall is considered "below normal" if it falls in the range of 90-96% of the LPA

IMD forecasts 'normal' monsoon

Vishwa.Mohan
@timesgroup.com

New Delhi: In what will spell a measure of relief for the economy reeling under the impact of the Covid-19 pandemic, the India Meteorological Department (IMD) on Wednesday predicted a 'normal' summer monsoon for the country during the June-September period and also released new reference dates for its onset or arrival and withdrawal in various cities.

The new dates indicate whether the monsoon would be longer or shorter by a few days or even a week in cities and states but the bigger and more reassuring message was that rainfall is likely to be 100% of the long period average, with a plus-minus error of 5%. This spells encouraging news for the rural economy as a good monsoon and kharif harvest will help sustain rural demand and s can keep rural

"Quantitat soon seasonal tember) rainfa 100% of the L erage with a m

5%," said Madhavan Rajeevan, secretary, ministry of earth sciences. Releasing the southwest monsoon forecast, he said there was quite a high probability of 2020 being a year of good rainfall (normal to excess). More than any other

FORECAST FOR 2020

- Southwest monsoon rainfall over the country as a whole is likely to be normal (96-104% of the Long Period Average)
- The LPA of the season (June-September) rainfall over the country as a whole for the period 1961-2010 is 88 cm

New dates for onset of monsoon

City	Old date	New date
Chennai	June 1	June 4
Pune	June 9	June 10
Mumbai	June 10	June 11
Kolkata	June	
Lucknow	June	
Delhi	June	
Chandigarh	July	

Source: The Times of India, Hyderabad edition, April 16, 2020

IMD predicted 65% of rainfall to be 'normal'

▶From P 1

The IMD's first stage forecast for the seasonal rainfall over the country as a whole, issued in April, was 100% of LPA with a model error of ± 5% of LPA. The forecast was, however, upgraded to 102% of LPA with a model error of ± 4% of LPA in the IMD's update in May end.

The IMD also predicted a probability of 65% of monsoon rainfall to be 'normal' to 'above normal'. The actual seasonal rainfall for the country as a whole was 109% of LPA, which is more than the predicted value and thus turned out to be very positive for the kharif (summer sown) crops.

The monsoon started retreating from western parts of north-west India on September 28 against the normal withdrawal date of September 17 with a delay of around 11 days. "Withdrawal has been delayed mainly due to active monsoon trough in association with the formation of two low pressure systems in September," said the IMD.

As on Thursday, the southwest monsoon has withdrawn from Punjab, western Himalayan region, Haryana, Chandigarh, Delhi and many parts for Rajasthan and some parts of Uttar Pradesh.

Source: The Times of India, Hyderabad edition, October 02, 2020

Statistics in Manufacturing

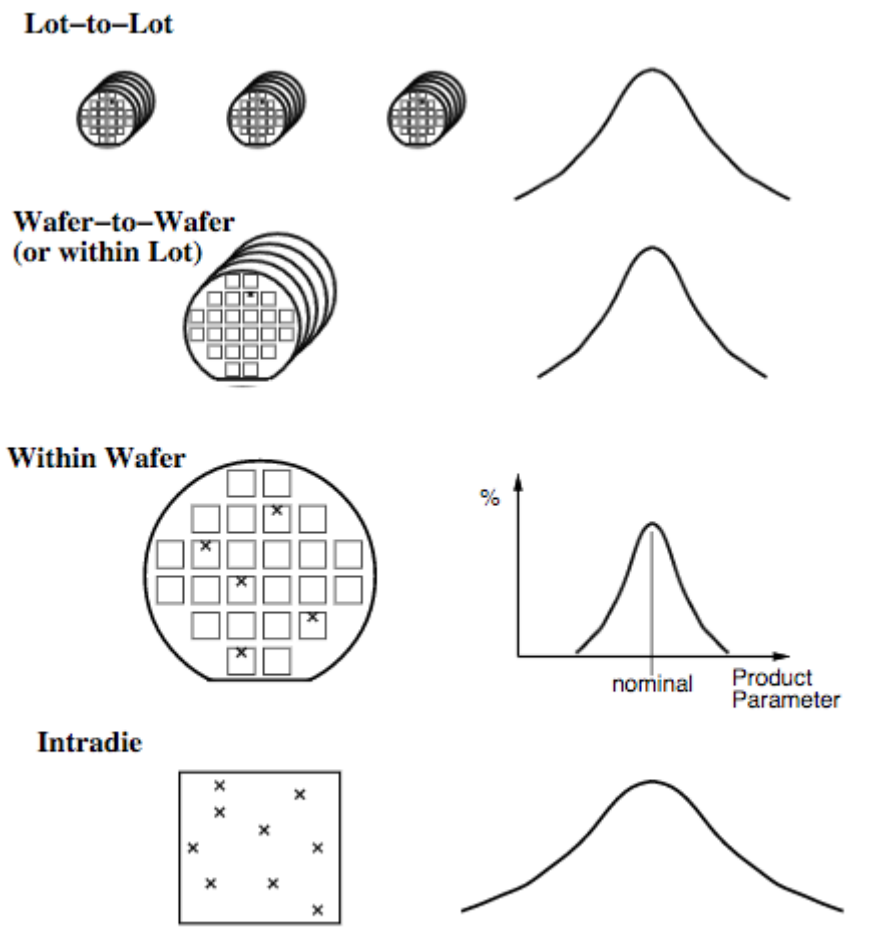
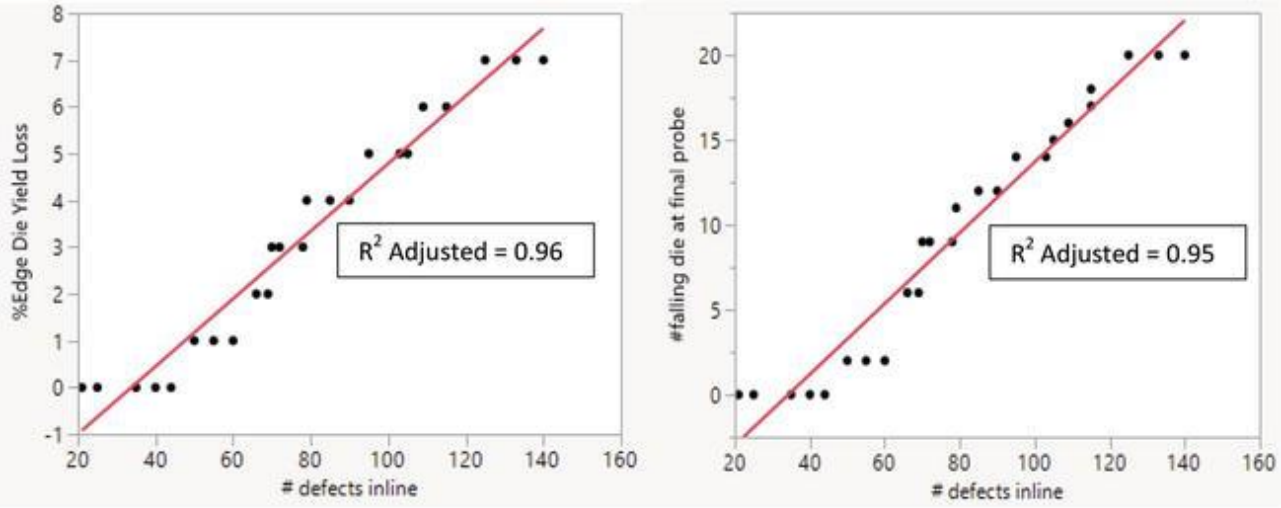


Figure 1. Spatial and temporal variation scales.



q*	Alpha
1.95996	0.05

Level	- Level	Score Mean Difference	Std Err Dif	Z	p-Value
Tool 3	Tool 2	-0.4615	2.995381	-0.15408	0.8775
Tool 2	Tool 1	-12.9231	2.998974	-4.30917	<.0001*
Tool 3	Tool 1	-12.9231	2.998974	-4.30917	<.0001*

Source: https://www.intechopen.com/books/applications-of-design-for-manufacturing-and-assembly/application-of-six-sigma-in-semiconductor-manufacturing-a-case-study-in-yield-improvement
Last accessed: April 17, 2020

Statistics in Social, Behavioural and Miscellaneous Studies

Pvt research plays down risk from mobiles: Study

DurgeshNandan.Jha
@timesgroup.com

New Delhi: Is radiation from mobile phones harmful? Multiple studies globally have not conclusively reached an answer. But an analysis by AIIMS of all research on the subject has found an interesting pattern — government funded studies show increased risk of brain tumour on long-term exposure to mobile phone radiation while industry-funded research tends to underestimate the risk.

"We found that industry-funded studies are not of good quality and tend to underestimate the risk. Government funded studies show increased risk of brain tumour on long-term exposure," said Dr Kameshwar Prasad, head of neurology at AIIMS, who is lead author of the study.

PHONE TROUBLE

- **1.33 times higher risk of brain tumour due to mobile use**, according to average of all studies
- Mixed-funded (govt, industry, mobile makers) research says **risk much lower, at 1.05 times**
- Studies funded by phone industry and mixed sources have **low quality score of 5-6**
- Score of govt-funded **research higher at 7-8**

▶Brain tumour risk, P 12

How hand sanitisers can harm children

New York: Scientists have warned that hand sanitisers might do more harm than good. They have found these alcohol-based, scented products might tempt young kids to swallow the substance — leading to stomach pain, nausea, apnea and even coma.

Researchers from US Centres for Disease Control (CDC) and Prevention have identified serious consequences, including apnea, acidosis and coma in young children who swallowed alcohol-based hand sanitisers.

To characterise paediatric alcohol hand sanitiser exposures, researchers from US Centres for Disease Control and Prevention analysed data reported by poison centres among children aged 12 years during 2011 to 2014.

The study found majority of intentional exposures to alcohol hand sanitisers occurred in children aged 6-12 years. During 2011-2014, 70,669 hand sanitiser exposures in children aged 12 years were reported, of which 65,293 (92%) were alcohol exposures and 5,376 (8%) were non alcohol exposures. These data also indicate that, among older children, exposures occur less frequently during the summer months. PTI

Women earned 25% less than men in 2016: Study

TIMES NEWS NETWORK

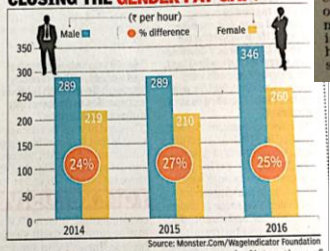
Mumbai: The average gender pay gap in India may have narrowed by two percentage points since last year, but it still stands at a yawning 25%. According to Monster Salary Index (MSI) data from 2016, men earned a median gross hourly salary of Rs 346, while women earned only Rs 260 in 2016. While it is an improvement over last year's over 27%, the pay gap is a notch higher than 24% in 2014. In the manufacturing sector, the average gender pay gap is as high as 30%.

Sanjay Modi, managing director (APAC & Middle East), Monster.com, said, "In India, the gender pay gap story holds true and the overall gap across India Inc is at 25%. This primarily is a manifestation of the underlying diversity challenges that organisations currently face."

According to the survey, which aimed at understanding the working women of India and their workplace concerns, nearly 44% women feel that gender parity need not be a top priority for their organisations. On the other hand, almost 69% expressed that even if gender parity is a priority, the management "does not walk the talk", indicating the need for India Inc to step up and implement pragmatic policies to bridge the gap.

"Unexamined conventions of women's commitment to work, distractions of family commitments, societal perception of women who work for long hours, etc, are the challenges that constrain women's progress. Sensitisation towards gender diversity and a step towards examining our socialised preconceptions would help break barriers to a more fair and inclusive work environment," said Modi.

CLOSING THE GENDER PAY GAP?



Source: Monster.Com/WageIndicator Foundation

Alcohol consumption can attract mosquitoes

DC CORRESPONDENT
HYDERABAD, JAN. 9

Alcohol consumption can attract mosquitoes as the body odour is strong, according to a study published by the American Mosquito Control Association.

The study found that those who had taken alcohol had a larger number of mosquitoes attracted to them. It was found that those who have beer attracts more mosquitoes because of the smell emitted from the body is stronger. It was found that drinking beer increased ethanol content in the sweat which lures mosquitoes. The other reason was intake of alcohol increased the body temperature and the heat makes the mosquito rush towards these people.

Entomologist Mr K. Venkatesh said, "Studies have shown that those who eat a lot of sugar and sweets have more and more mosquitoes attracted to them. It is the body odour which attracts the mosquitoes. The alcohol is mostly fermented, hence the body odour is strong."

- **MOSQUITOES ARE** attracted towards those who consume alcohol as their body temperature rises
- **AMONGST ALCOHOL** drinkers, beer seems to be the favourite of mosquitoes
- **PEOPLE WHO** drink alcohol are susceptible to mosquito-borne diseases
- **MOSQUITOES** also look at body type to get attracted.

Researchers stated that their experiments showed that those who consumed alcohol had mosquitoes going towards them. This means that the alcohol drinkers will be the first target of mosquitoes.

The increasing burden of mosquito-borne diseases like malaria, dengue and zika is making researchers study body odour which attracts mosquitoes.

Dr G H Kishan, senior general physician, added, "The cities have become breeding ground for mosquito. To add to this, there are a good number of human bodies that create the right odour to attract them. Due to this reason, the spread is large and wide in urban areas."

CLIMATE HAS A NOSE FOR SHAPES

THE SHAPE OF OUR NOSES WAS FORMED BY A LONG PROCESS OF ADAPTATION TO CLIMATE, A STUDY SAYS

- **WIDER** noses are more common in warm-humid climates
- **NARROWER** noses are more common in cold-dry climates
- **NARROWER** nasal passages help to increase the moisture content of air and warm it — a bonus for those in higher latitudes

EVOLUTION OF THE NOSE

The width of the nostrils and the base of the nose measurements differ across populations, researchers found

This indicates a role for natural selection in the evolution of nose shape.

The width of the nostrils is strongly correlated with temperature and absolute humidity

ADAPTING MECHANISM

Narrower nostrils alter the airflow so the mucous-covered inside of the nose can humidify and warm the air

People with narrower nostrils therefore fared better in colder climates

This lead to a gradual decrease in nose width in populations living far away from the equator

Half of all pregnant women are anaemic

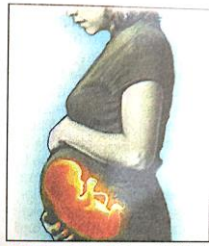
▶Continued from P 1

The WHO defines wasting as low weight for height, stunting as low height for age, and underweight as low weight for age.

The survey also found that just over half of all pregnant women were anaemic. This would automatically translate into their newborn being weak. Overall, 53% of women and 23% of men in the 15-49 age group were anaemic.

There is wide variation among states. The data for UP has not been released in view of the ongoing polls, according to Balram Paswan, professor at Mumbai-based International Institute for Population Sciences which was the nodal agency for the survey done for the health ministry. But poorer states like Bihar, Madhya Pradesh, Jharkhand, Assam, Rajasthan and Chhattisgarh have higher than national average rates on all markers.

More advanced states like those in the south, Haryana and Gujarat have slightly better numbers but are still at unacceptable levels. In Tamil Nadu, 51% children are anaemic while in Kerala it is over one-third. In many states, stunting has declined but the share of severely wasted children has increased. These are clear signs of an endemic crisis of hunger in the country that policy makers don't appear to be addressing.



HUNGER CRISIS: Overall, 53% of women and 23% of men in the 15-49 age group were anaemic

DSC 7215

Statistics in Social, Behavioural and Miscellaneous Studies

HEALTH | MATTERS ■ Heart patients advised to have seven minute bout of light physical activity

Move every 20 minutes, live long: Study

Toronto, Oct. 22: Heart patients should interrupt sedentary time every 20 minutes with a seven minute bout of light physical activity to prolong life, according to a study.

Previous research has shown that being sedentary for long periods could shorten life but taking breaks to move around may counteract the risk, particularly if it means burning more than 770 kilo calories (kcal) a day.

This study, presented at the annual Canadian Cardiovascular Congress (CCC) in Toronto, investigated how many breaks, and for what duration, are needed to expend 770 kcal.

"Our study shows that heart patients should interrupt sedentary time every 20 minutes with a seven minute bout of light physical activity," said Ailar Ramadi from the University of Alberta in Canada.

"Simple activities such as standing up and walking at a casual pace will expend more than 770 kcal in a day if done with this frequency and duration," Ramadi said.

■ Being sedentary for long periods could shorten life but taking breaks may counteract risk.

■ The study was presented at the annual Canadian Cardiovascular Congress in Toronto.

■ There is a lot of evidence now that sitting for long periods is bad for health.

The study enrolled 132 patients with coronary artery disease. The average age was 63 years and 77 per cent were male. Participants wore an armband activity monitor for an average of 22 hours a day for five days.

The activity monitor recorded the amount of energy spent during breaks from inactivity, the amount of inactive time, and the number and duration of breaks during each sedentary hour.

"There is a lot of evidence now that sitting for long periods is bad for health. Our study suggests that during each hour of sitting time, heart patients should take three breaks which add up to 21 minutes of light physical activity, said Ramadi.

"This will expend 770 kcal a day, an amount associated with a lower risk of premature death," Ramadi said.

– PTI

WAKE-UP | CALL ■ Experiment tries to link sleep loss and feeling negative in that moment

Lack of sleep makes people angry, study finds

Washington, Nov. 28: Losing just a couple hours of sleep at night makes you angrier, especially in frustrating situations, a study has found.

Other studies have shown a link between sleep and anger, but questions remained about whether sleep loss was to blame or if anger was responsible for disrupted sleep, said Zlatan Krizan, a professor at Iowa State University in the US.

The research, published in the *Journal of Experimental Psychology: General*, answers those questions and provides new insight on our ability to adjust to irritating conditions when tired.

"Despite typical tendencies to get somewhat used to irritating conditions — an uncomfortable shirt or a barking dog — sleep-restricted individuals actually showed a trend toward increased anger and distress, essentially reversing their ability to adapt to frustrating conditions over time. No one has shown this before," Krizan said.

Study participants were randomly split into two groups: one maintained their normal sleep routine and the second restricted their sleep by two to four hours each night for two nights.

SLEEP MANIPULATION

■ Participants in a sleep manipulation project were tested in a lab

■ They were subjected to rate different products while listening to two different noises

■ Those who had not slept enough reacted more angrily under annoying noise

■ Effects measured to understand relationship between sleep and emotions

Those who maintained averaged almost seven hours of sleep a night, while the restricted group got about four and a half hours each night.

The difference reflects sleep loss we regularly experience in everyday life, Krizan said.

To measure anger, researchers had participants come to the lab — before and after the sleep manipulation — to rate different products while listening to brown noise (similar to the sound of spraying water) or more aversive white noise (similar to a static signal).

Krizan said the purpose was to create uncomfortable conditions, which tend to provoke anger.

"In general, anger was substantially higher for those who were sleep restricted. We manipulated how annoying the noise was during the task and as expected, people reported more anger when the noise was more unpleasant," said Krizan.

"When sleep was restricted, people reported even more anger, regardless of the noise," he said.

It is well established that sleep loss increases negative emotions like anxiety and sadness, and decreases positive emotions, such as happiness and enthusiasm, Krizan said.

Researchers measured these effects to more generally understand the relationship between sleep, anger and emotions. Krizan said they found sleep loss to uniquely impact anger, and not just result from feeling more negative in that moment.

They also tested whether subjective sleepiness explained more intense feelings of anger.

— PTI

FRESH | FINDINGS ■ Research says it can improve concentration and multitasking abilities

Spinning charkha daily could improve behaviour

New Delhi, Nov. 20: Spinning charkha for one hour daily may improve concentration, multitasking abilities, patience and mind-body coordination of children and youth, a study claims.

The study conducted by the Brain Behaviour Research Foundation Of India (BBRFI) analysed the relevance of charkha (spinning wheel) among children and youth in the society.

According to the research conducted for over a month on 30 individuals, 10 to 18 years of age, spinning a charkha everyday had positive and soothing effects on the participants.

Pre and post training data was collected using brain mapping & electroencephalogram (EEG). Subjects were also tested for psychological, personality and intelligence quotient (IQ) tests.

The study saw positive results in all aspects of behaviour, according to the research recently presented by Chairperson of BBRFI, Meena Mishra.

It provides scientific foundation to anecdotal reports of benefits of charkha spinning, Mishra said. The study was sponsored by Gandhi Darshan, Ministry of Culture.

The magazine shines the spotlight on issues, approaches and prevailing commentary on mental health in India and across the world. "Mental health is even more important that physical health. At AIIMS we are actively looking at working with our psychology department to help family members of patients cope with some trying and difficult times," said Rajesh Malhotra from All India Institute of Medical Sciences, New Delhi.

Commenting on the status of mental health in India, Rajesh Sagar, Professor, Department Of Psychiatry, AIIMS, Delhi, said, "There is a great deal of stigma in the society on mental health. Mental wellbeing is often construed as being crazy or mad, which mean people find it difficult to talk about and even seek help in intervention. We, as a society, need to change this attitude."

The urgency of focus on mental health in India can be pointed out with the fact that, one in four people in the world is likely to be affected by mental or neurological disorders at some point in their lives, the researchers said.

—PTI

Women more likely to marry men with firm handshake

Chelsea Ritschel

A firm handshake is an important attribute if you are looking to settle down — as it turns women are more likely to marry men with a strong handshake, according to new research.

Researchers from Columbia University's Mailman School of Public Health and the Columbia Ageing Centre discovered the link between a firm grip and a higher likelihood of marriage after studying 5,000 adults from the Norwegian City of Tromso.

Thinkstock

GET A GRIP

According to the researchers, males with a stronger grip were more likely to be married compared to their flimsy-handed counterparts. However, the same was not true for women. The study, published in the journal SSM-Population Health, analysed the grip strength of two groups of people, born between 1923-1935 and 1936-1948 and aged 59 to 71. Each participant's grip strength was assessed using a vigorimeter — a device that asks participants to squeeze a rubber balloon and compared to the others in the group. According to the findings, women consider grip strength a marital benefit — as researchers found unmarried men were more likely to have low grip strength.

Vegard Skirbekk, a professor at the Columbia Ageing Centre and an author of the study said: "Our results hint that women may be favouring partners who signal strength and vigour when they marry.

"If longer-lived women marry healthier men, then both may avoid or defer the role of caregiver, while less healthy men remain unmarried and must look elsewhere for assistance." In addition to a higher likelihood of tying the knot, a strong grip is an indicator of a person's capacity to be socially active and healthy — and of heart health.

According to Bhargav Hanagal, ... Medical School, grip strength is a useful tool in determining an individual's risk of heart attack or stroke as it suggests overall strength. The method is so helpful that it is a "better predictor of death or cardiovascular disease than blood pressure." And researchers at the University of Manchester have found that a strong grip is often a sign of better problem-solving skills, memory, and fast reaction times. So the next time you consider a potential suitor, don't forget to shake their hand. INDEPENDENT

National Agricultural Cooperative Marketing

DSC 7215

Why Study Statistics?

Statistics are part of our daily life and are all around us…

…They are also used and misused heavily…

…and so, we must study and understand Statistics.

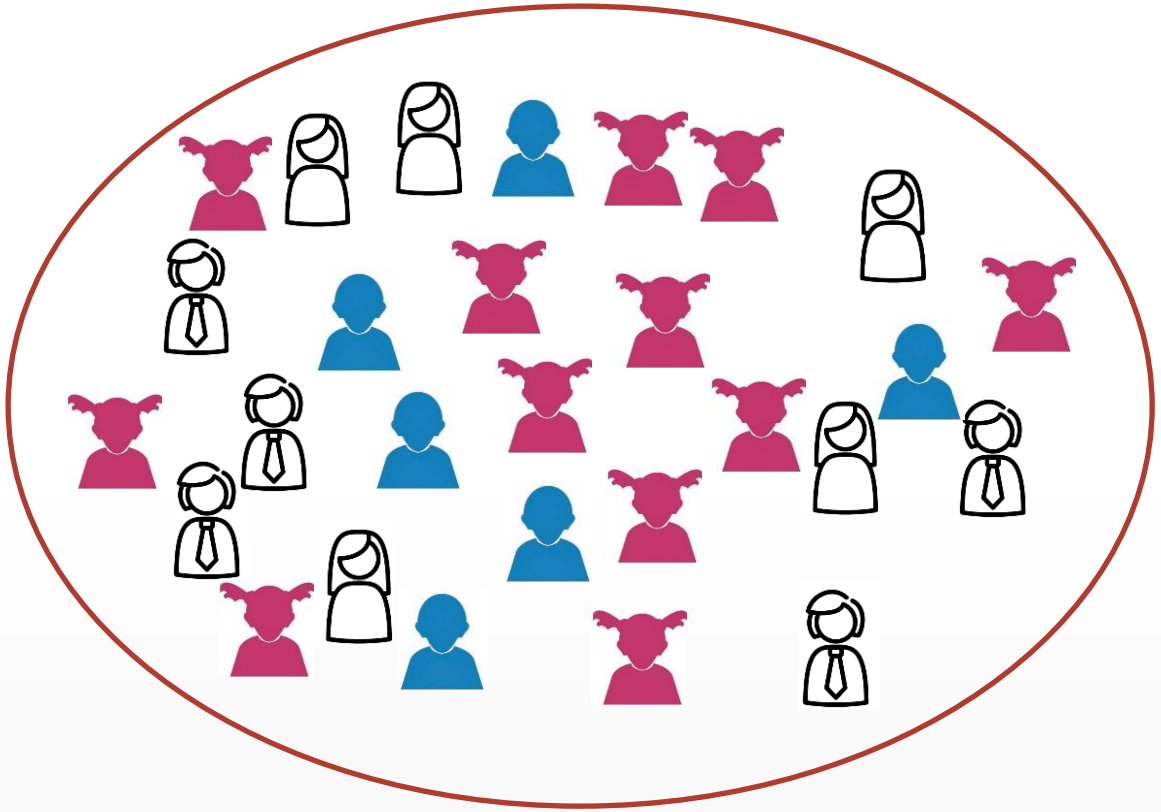
They are also the foundation for Data Science.

DSC 7215

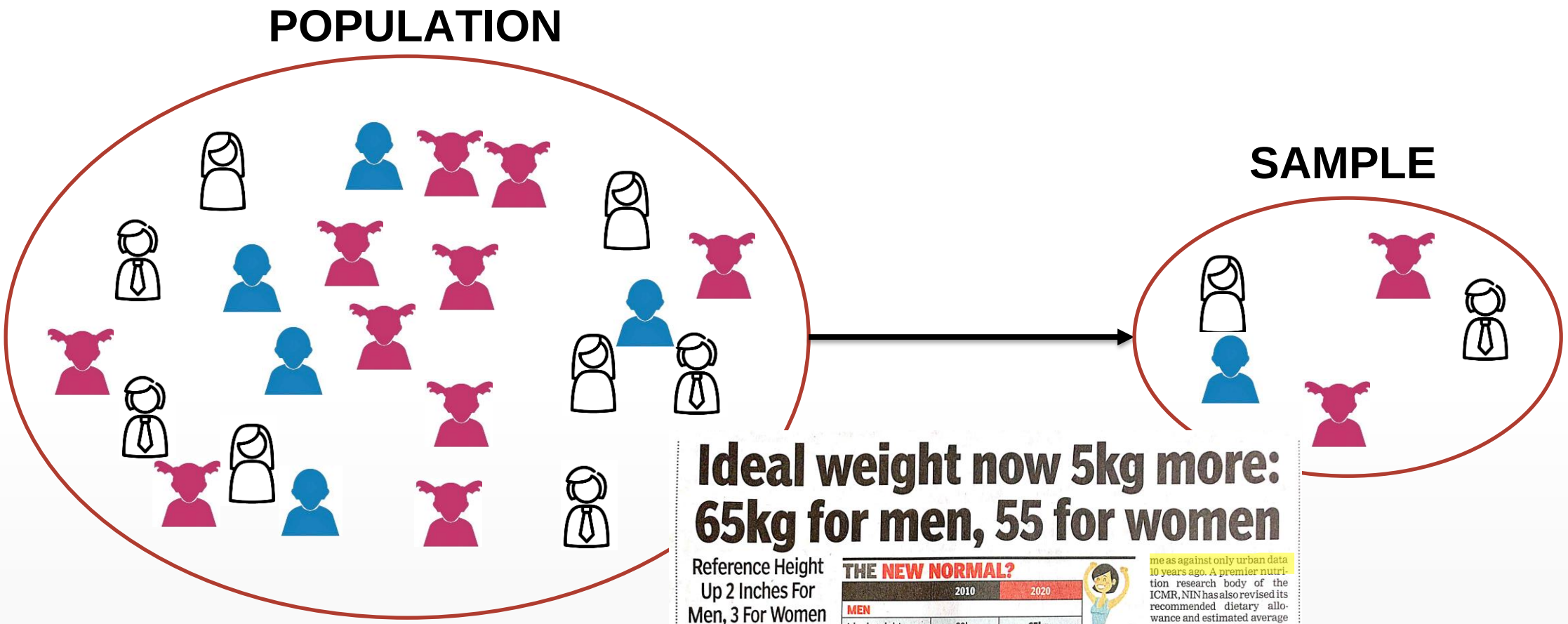
Basic Statistical Terminology

Population and Sample

POPULATION



Population and Sample



Ideal weight now 5kg more: 65kg for men, 55 for women

Reference Height Up 2 Inches For Men, 3 For Women

Syed.Akbar@timesgroup.com

Hyderabad: The National Institute of Nutrition (NIN) has added another 5kg to the ideal weight of Indians. While the weight of an ideal or reference Indian man was 60kg in 2010, it has now increased to 65kg. In case of women, it has gone up to 55kg from 50kg a decade ago.

Along with this, the earlier height for a reference Indian man was 5.6 feet (171cm) and a woman was 5 feet (152cm). It is now revised to 5.8 feet (177cm) for a man and 5.3 feet (162cm) for a woman. This will now be taken for normal body mass index.

The scientists said this was because nutritional food intake had gone up. Also, data from rural and urban areas was taken into account this time as against only urban data 10 years ago. A premier nutrition research body of the ICMR, NIN has also revised its recommended dietary allowance and estimated average requirement of nutrition for Indians in its 2020 report, released on Monday. The definition for reference Indian adult man and woman with regard to age was changed to 19-39 years instead of 20-39 years, which was fixed in 2010. The scientists added that a previous expert committee in 1989 had taken the body weight and height of only well-to-do Indian children and adolescents.

THE NEW NORMAL?

	2010	2020
MEN		
Ideal weight	60kg	65kg
Reference height	5.6ft (171cm)	5.8 ft (177cm)
WOMEN		
Ideal weight	50kg	55kg
Reference height	5ft (152 cm)	5.3ft (162 cm)

Recommended common salt intake 5g/day with sodium limited to 2g/per day. Intake of 3,510mg potassium recommended/day

▶Data from all states: P6

Also read: The 1936 *Literary Digest* Poll
https://www2.math.upenn.edu/~deturck/m170/wk4/lecture/case1.html

Parameter and Statistic

Parameter: A descriptive measure of the population.

For example, population mean, population variance, population standard deviation, etc.

Statistic: A descriptive measure of the sample.

For example, sample mean, sample variance, sample standard deviation, etc.

"For the 1936 election, the Literary Digest prediction was that Landon would get 57% of the vote against Roosevelt's 43% (these are the statistics that the poll measured). The actual results of the election were 62% for Roosevelt against 38% for Landon (these were the parameters the poll was trying to measure). The sampling error in the Literary Digest poll was a whopping 19%, the largest ever in a major public opinion poll. Practically all of the sampling error was the result of sample bias."
https://www2.math.upenn.edu/~deturck/m170/wk4/lecture/case1.html

Parameter and Statistic

<u>Greek</u> – Population **Parameter**
Mean – μ
Variance – σ^2
Standard Deviation - σ

<u>Roman</u> – Sample **Statistic**
Mean – $\bar{x}$
Variance – s^2
Standard Deviation - s

Descriptive and Inferential Statistics

- **Descriptive Statistics** – Data gathered about a group (sample or population) to reach conclusions about the <u>same</u> group.
- **Inferential Statistics** – <u>Data gathered from a sample</u> and the statistics generated to reach <u>conclusions about the population</u> from which the sample is taken.

1 Diabetes is a huge problem in India.

The prevalence of diabetes increased tenfold, from 1.2% to 12.1%, between 1971 and 2000.
Noncommunicable Diseases in the Southeast Asia Region, Situation and Response, World Health Organization, 2011. http://apps.searo.who.int/PDS_DOCS/B4793.pdf

It is estimated that 61.3 million people aged 20-79 years live with diabetes in India (2011 estimates). This number is expected to increase to 101.2 million by 2030.
David R. Whiting, et al. IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030, Diabetes Research and Clinical Practice, Volume 94, Issue 3, December 2011, Pages 311-321, http://www.sciencedirect.com/science/article/pii/S0168822711005912)

And, 77.2 million people in India are said to have pre-diabetes.
Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, et al. "Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: phase I results of the Indian Council of Medical Research-India Diabetes (ICMR-INDIAB) study" Diabetologia 54.12 (2011): 3022-7. NCBI. Web. March 2013.

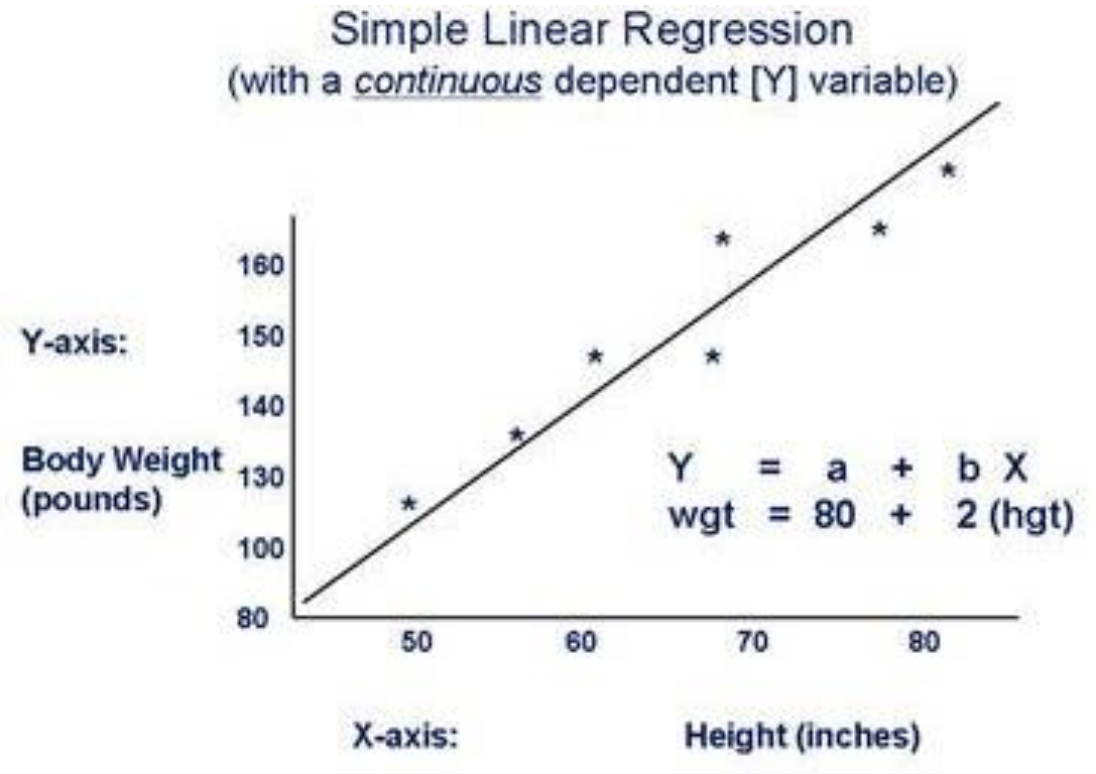
Source: http://www.arogyaworld.org/wp-content/uploads/2010/10/ArogyaWorld_IndiaDiabetes_FactSheets_CGI2013_web.pdf
Last accessed: July 22, 2021

Variables and Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
35	management	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1	0	unknown	no
28	management	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1	0	unknown	no
42	entrepreneur	divorced	tertiary	yes	2	yes	no	unknown	5	may	380	1	-1	0	unknown	no
58	retired	married	primary	no	121	yes	no	unknown	5	may	50	1	-1	0	unknown	no
43	technician	single	secondary	no	593	yes	no	unknown	5	may	55	1	-1	0	unknown	no
41	admin.	divorced	secondary	no	270	yes	no	unknown	5	may	222	1	-1	0	unknown	no
29	admin.	single	secondary	no	390	yes	no	unknown	5	may	137	1	-1	0	unknown	no
53	technician	married	secondary	no	6	yes	no	unknown	5	may	517	1	-1	0	unknown	no
58	technician	married	unknown	no	71	yes	no	unknown	5	may	71	1	-1	0	unknown	no
57	services	married	secondary	no	162	yes	no	unknown	5	may	174	1	-1	0	unknown	no
51	retired	married	primary	no	229	yes	no	unknown	5	may	353	1	-1	0	unknown	no
45	admin.	single	unknown	no	13	yes	no	unknown	5	may	98	1	-1	0	unknown	no
57	blue-collar	married	primary	no	52	yes	no	unknown	5	may	38	1	-1	0	unknown	no
60	retired	married	primary	no	60	yes	no	unknown	5	may	219	1	-1	0	unknown	no
33	services	married	secondary	no	0	yes	no	unknown	5	may	54	1	-1	0	unknown	no
28	blue-collar	married	secondary	no	723	yes	yes	unknown	5	may	262	1	-1	0	unknown	no
56	management	married	tertiary	no	779	yes	no	unknown	5	may	164	1	-1	0	unknown	no
32	blue-collar	single	primary	no	23	yes	yes	unknown	5	may	160	1	-1	0	unknown	no
25	services	married	secondary	no	50	yes	no	unknown	5	may	342	1	-1	0	unknown	no
40	retired	married	primary	no	0	yes	yes	unknown	5	may	181	1	-1	0	unknown	no

Variables – Dependent and Independent

- Dependent variables on y-axis and Independent on x-axis.
- Dependent variable also called Target variable or Class variable.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
35	management	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1	0	unknown	no
28	management	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1	0	unknown	no
42	entrepreneur	divorced	tertiary	yes	2	yes	no	unknown	5	may	380	1	-1	0	unknown	no
58	retired	married	primary	no	121	yes	no	unknown	5	may	50	1	-1	0	unknown	no
43	technician	single	secondary	no	593	yes	no	unknown	5	may	55	1	-1	0	unknown	no
41	admin.	divorced	secondary	no	270	yes	no	unknown	5	may	222	1	-1	0	unknown	no
29	admin.	single	secondary	no	390	yes	no	unknown	5	may	137	1	-1	0	unknown	no
53	technician	married	secondary	no	6	yes	no	unknown	5	may	517	1	-1	0	unknown	no
58	technician	married	unknown	no	71	yes	no	unknown	5	may	71	1	-1	0	unknown	no
57	services	married	secondary	no	162	yes	no	unknown	5	may	174	1	-1	0	unknown	no
51	retired	married	primary	no	229	yes	no	unknown	5	may	353	1	-1	0	unknown	no
45	admin.	single	unknown	no	13	yes	no	unknown	5	may	98	1	-1	0	unknown	no
57	blue-collar	married	primary	no	52	yes	no	unknown	5	may	38	1	-1	0	unknown	no
60	retired	married	primary	no	60	yes	no	unknown	5	may	219	1	-1	0	unknown	no
33	services	married	secondary	no	0	yes	no	unknown	5	may	54	1	-1	0	unknown	no
28	blue-collar	married	secondary	no	723	yes	yes	unknown	5	may	262	1	-1	0	unknown	no
56	management	married	tertiary	no	779	yes	no	unknown	5	may	164	1	-1	0	unknown	no
32	blue-collar	single	primary	no	23	yes	yes	unknown	5	may	160	1	-1	0	unknown	no
25	services	married	secondary	no	50	yes	no	unknown	5	may	342	1	-1	0	unknown	no
40	retired	married	primary	no	0	yes	yes	unknown	5	may	181	1	-1	0	unknown	no

Data – Numeric and Categorical

18 kg

27 kg

Sources: http://banglanews24.com/en/files/2013August/SM/Gold-sm20130830024804.jpg, http://myoor.com/wp-content/uploads/2014/01/gold.jpg and http://im.rediff.com/cricket/2014/feb/01india1.jpg
Last accessed: November 22, 2014

Categorical Data (Qualitative)

Nominal

Examples

- Employee ID
- Gender
- Religion
- Ethnicity
- Pin codes
- Place of birth
- Aadhaar numbers

Ordinal

Examples

- Mutual fund risk ratings
- Fortune 50 rankings
- Movie ratings

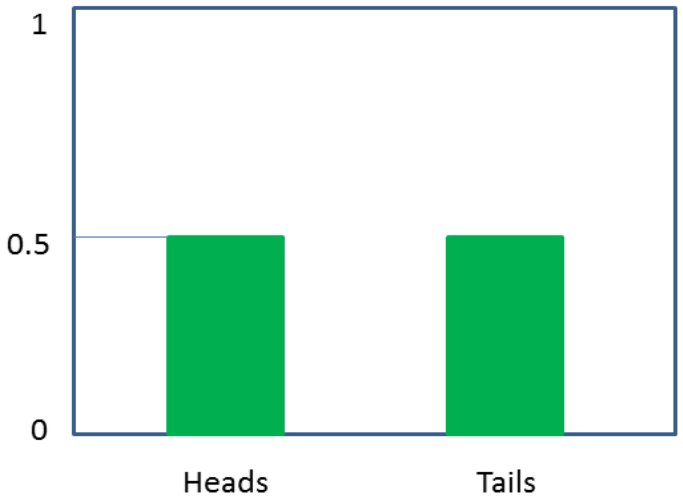
While there is an order, difference between consecutive levels are not always equal.

Numeric Data (Quantitative)

Examples

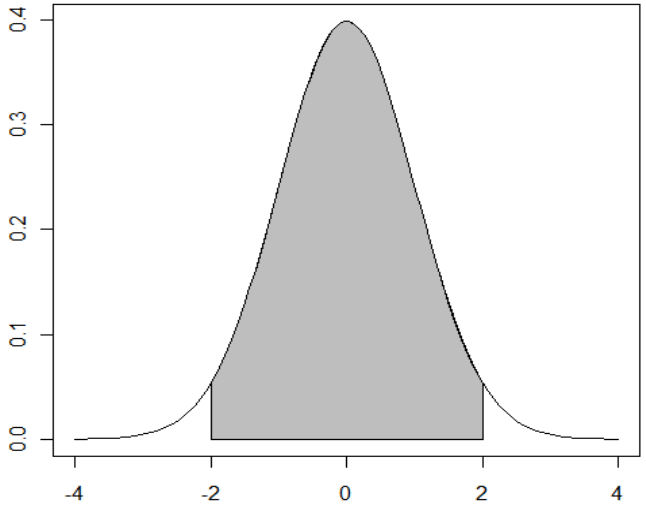
- Number of iPads sold
- Number of complaints received at the call centre
- Number of employees
- Height
- Weight
- Time
- Volume
- Percentage return on a stock
- Rupee change in stock price

Numeric - Discrete and Continuous



Countable

In any given range, only certain values are POSSIBLE.



Measurable

In any given range, all values are POSSIBLE.

DSC 7215

Discrete or Continuous?

Time between customer arrivals at a retail outlet	Continuous
No. of customers arriving at a retail outlet during a five-minute period	Discrete
Sampling 100 voters in an exit poll and determining how many voted for the winning candidate	Discrete
No. of new coronavirus cases per day in India	Discrete
Time taken for doubling of coronavirus cases in cities around the world	Continuous

Describing Data Through Statistics

Central Tendencies

Demographic characteristics

Among 55,924 laboratory confirmed cases reported as of 20 February 2020, the median age is 51 years (range 2 days-100 years old; IQR 39-63 years old) with the majority of cases (77.8%) aged between 30–69 years. Among reported cases, 51.1% are male, 77.0% are from Hubei and 21.6% are farmers or laborers by occupation.

Below is an extract of the study findings (highlight added by Worldometer):

Among the first 425 patients with confirmed NCIP, the median age was 59 years and 56% were male. The majority of cases (55%) with onset before January 1, 2020, were linked to the Huanan Seafood Wholesale Market, as compared with 8.6% of the subsequent cases.

The mean incubation period was 5.2 days (95% confidence interval [CI], 4.1 to 7.0), with the 95th percentile of the distribution at 12.5 days.

In its early stages, the epidemic doubled in size every 7.4 days. With a mean serial interval of 7.5 days (95% CI, 5.3 to 19), the basic reproductive number was estimated to be 2.2 (95% CI, 1.4 to 3.9).

Real-World Decision Making

Calculating **descriptive statistics** to get an understanding of the data, without which no modeling can be effective.

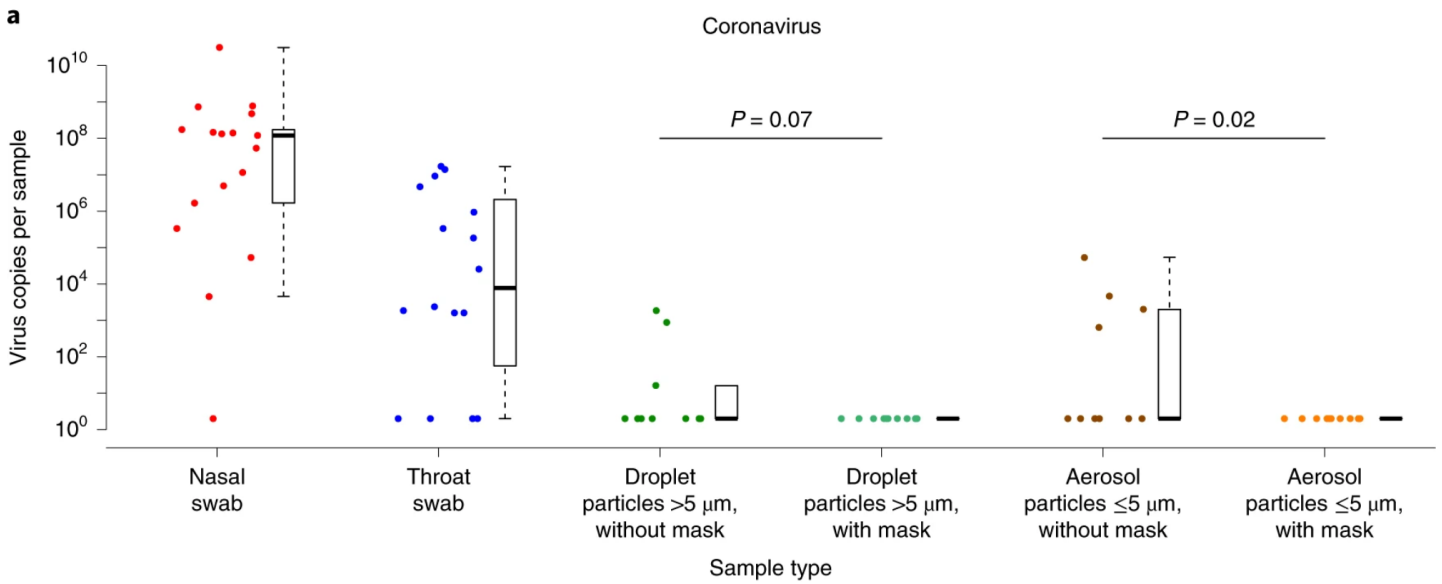
Source: Report of the WHO-China Joint Report on Coronavirus Disease 2019 (COVID-19), 16-24 February, 2020 and https://www.worldometers.info/coronavirus/coronavirus-incubation-period/#asy
Last accessed: March 20, 2020

Fig. 1: Efficacy of surgical face masks in reducing respiratory virus shedding in respiratory droplets and aerosols of symptomatic individuals with coronavirus, influenza virus or rhinovirus infection.

From: Respiratory virus shedding in exhaled breath and efficacy of face masks

Real-World Decision Making

More **descriptive statistics** and then modeling to determine if face masks are effective in reducing virus shedding.



differences in bold. For nasal swabs and throat swabs, all infected individuals were included (coronavirus, n=17; influenza virus, n=43; rhinovirus, n=54). For respiratory droplets and aerosols, numbers of infected individuals who provided exhaled breath samples while not wearing or wearing a surgical face mask, respectively were: coronavirus (n=10 and 11), influenza virus (n=23 and 28) and rhinovirus (n=36 and 32). A subset of participants provided exhaled breath samples for both mask interventions (coronavirus, n=4; influenza virus, n=8; rhinovirus, n=14). The box plots indicate the median with the interquartile range (lower and upper hinge) and ±1.5×interquartile range from the first and third quartile (lower and upper whiskers).

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020) https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

Understanding Data – Respiratory Virus Shedding in Exhaled Breath and Efficacy of Face Masks

Participant ID	Age	Female	No. of Coughs During Exhaled Breath Collection in 30 min	With Mask	PCR Viral (Coronavirus) Load (copies per ml) in Nasal Swab	PCR Viral (Coronavirus) Load (copies per ml) in Throat Swab
B10176	26	0	2	0	292000000.000	1860000.000
B10206	24	1	1	0	4622481.445	638.725
B10206	24	1	6	1	4622481.445	638.725
B10245	41	0	4	1	NA	NA
B10247	23	1	3	0	1814.573	NA
B10247	23	1	0	1	1814.573	NA
B10250	58	1	99	1	21400.000	NA
B10253	66	1	24	0	21400000.000	NA
B10258	30	1	69	1	135000.000	NA
B10260	63	1	0	1	12500000000.000	6750000.000
B10277	24	0	8	1	191000000.000	73500.000
B10284	32	1	11	1	53200000.000	135000.000
B10299	26	1	88	0	58970346.880	3674010.313
B10303	46	1	18	0	47905000.000	10230.000
B10313	35	1	4	1	56148406.880	637.166
B10315	52	1	2	0	1990314.648	5518837.500
B10315	52	1	3	1	1990314.648	5518837.500
B10360	28	1	5	0	69300000.000	748.000
B10394	26	0	0	1	307050342.500	377729.714
B10398	59	0	0	0	669032.515	NA
B10398	59	0	2	1	669032.515	NA

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

Descriptive Statistics

Respiratory Virus Shedding in Exhaled Breath and Efficacy of Face Masks

rhinovirus to have a measured fever. Coronavirus-infected participants coughed the most with an average of 17 (s.d. = 30) coughs during the 30-min exhaled breath collection. The profiles of the participants randomized to with-mask versus without-mask groups were similar

Demographic characteristics

Among 55,924 laboratory confirmed cases reported as of 20 February 2020, the median age is 51 years (range 2 days-100 years old; IQR 39-63 years old) with the majority of cases (77.8%) aged between 30–69 years. Among reported cases, 51.1% are male, 77.0% are from Hubei and 21.6% are farmers or laborers by occupation.

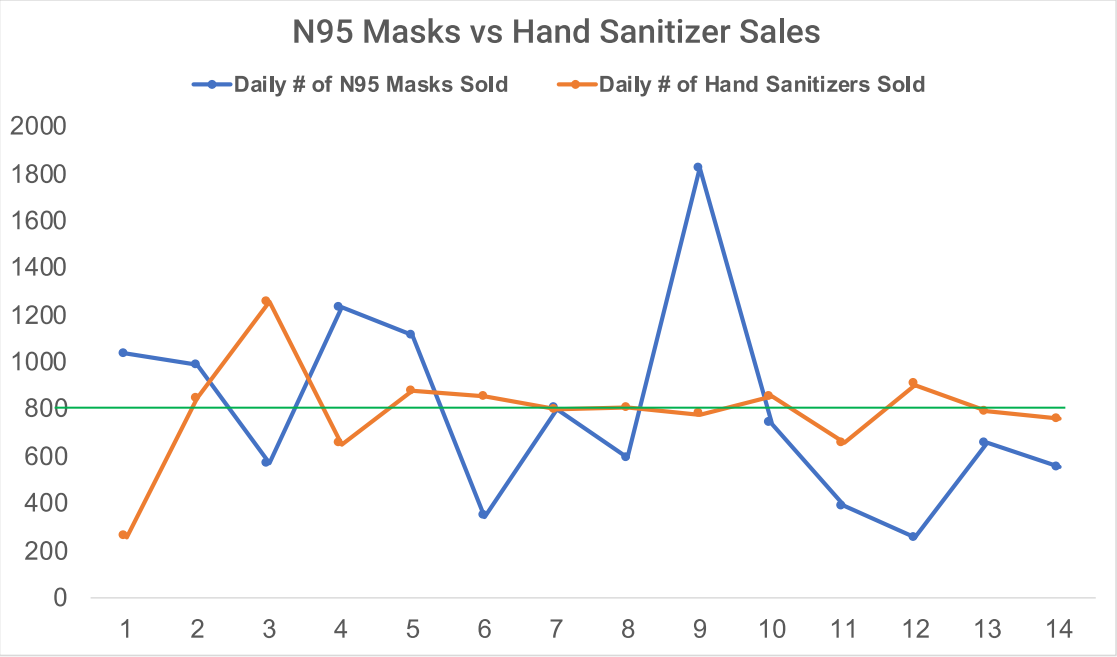
	All who provided exhaled breath	Coronavirus	Influenza virus	Rhinovirus
	(*n* = 246)	(*n* = 17)	(*n* = 43)	(*n* = 54)
	n (%)	*n* (%)	*n* (%)	*n* (%)
Female	144 (59)	13 (76)	22 (51)	30 (56)
Age group, years				
11–17	12 (5)	0 (0)	8 (19)	4 (7)
18–34	114 (46)	10 (59)	11 (26)	24 (44)
35–50	79 (32)	2 (12)	16 (37)	18 (33)
51–64	35 (14)	4 (24)	8 (19)	5 (9)
≥ 65	6 (2)	1 (6)	0 (0)	3 (6)
Chronic medical conditions				
Any	49 (20)	5 (29)	5 (12)	10 (19)
Respiratory	18 (7)	0 (0)	4 (9)	3 (6)
Influenza vaccination				
Ever	94 (38)	6 (35)	15 (35)	20 (37)
Current season	23 (9)	2 (12)	1 (2)	4 (7)
Previous season only	71 (29)	4 (24)	14 (33)	16 (30)
Ever smoker	31 (13)	1 (6)	6 (14)	6 (11)
Time since illness onset, h				
<24	22 (9)	0 (0)	5 (12)	2 (4)
24–48	100 (41)	9 (53)	13 (30)	25 (46)
48–72	85 (35)	8 (47)	18 (42)	20 (37)
72–96	39 (16)	0 (0)	7 (16)	7 (13)
History of measured fever ≥37.8 °C	58 (24)	3 (18)	17 (40)	8 (15)
Measured fever ≥37.8 °C at presentation	36 (15)	2 (12)	18 (42)	2 (4)
Measured body temperature (˚C) at enrollment (mean, s.d.)	36.8 (0.8)	36.9 (0.8)	37.4 (0.9)	36.6 (0.7)
Symptoms at presentation				
Fever	111 (45)	10 (59)	27 (63)	16 (30)
Cough	198 (80)	15 (88)	40 (93)	44 (81)
Sore throat	211 (86)	15 (88)	31 (72)	49 (91)
Runny nose	200 (81)	17 (100)	36 (84)	48 (89)
Headache	186 (76)	13 (76)	30 (70)	38 (70)
Myalgia	176 (72)	12 (71)	31 (72)	34 (63)
Phlegm	176 (72)	9 (53)	34 (79)	41 (76)
Chest tightness	64 (26)	3 (18)	12 (28)	9 (17)
Shortness of breath	103 (42)	6 (35)	14 (33)	25 (46)
Chills	100 (41)	8 (47)	29 (67)	16 (30)
Sweating	95 (39)	5 (29)	18 (42)	20 (37)
Fatigue	218 (89)	16 (94)	38 (88)	48 (89)
Vomiting	19 (8)	2 (12)	5 (12)	2 (4)
Diarrhea	17 (7)	2 (12)	1 (2)	6 (11)
Number of coughs during exhaled breath collection (mean, s.d.)	8 (14)	17 (30)	8 (11)	5 (9)

Sources: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2 and Report of the WHO-China Joint Report on Coronavirus Disease 2019 (COVID-19), 16-24 February, 2020 and https://www.worldometers.info/coronavirus/coronavirus-incubation-period/#asy
Last accessed: April 08, 2020

Descriptive Statistics

Summarize the data by understanding:

1. What is its **average** value – mean, median, mode **(central tendencies)**. These represent the most common value of a variable.
2. How far are data points **spread** – range, IQR (Interquartile Range), variance, standard deviation
3. How are they **distributed** – Box Plot



Daily # of N95 Masks Sold	Daily # of Hand Sanitizers Sold	
1034	259	
987	845	
567	1256	
1230	653	
1111	876	
345	854	
802	799	
591	804	
1823	777	
740	853	
389	657	
254	903	
657	789	
555	760	
11085	**11085**	**Sum**
791.8	**791.8**	**Mean**

- The average (mean) number of sales per day is the same for both N95 masks and hand sanitizers.
- However, daily sales of N95 masks fluctuate and are spread a lot more around the mean than the daily sales of hand sanitizers.

DSC 7215

Let Us Compute the Central Tendencies - Mean

	All who provided exhaled breath	Coronavirus	Influenza virus	Rhinovirus
	(*n* = 246)	(*n* = 17)	(*n* = 43)	(*n* = 54)
Number of coughs during exhaled breath collection (mean, s.d.)	8 (14)	17 (30)	8 (11)	5 (9)

$$\text{Mean, } \mu = \frac{\Sigma x}{n} = \frac{349}{21} = 16.6 \cong 17$$

Of the 17 people, 4 people provided cough samples both with and without mask, making the total records 21.

Participant ID	Age	Female	No. of Coughs During Exhaled Breath Collection in 30 min	With Mask
B10176	26	0	2	0
B10206	24	1	1	0
B10206	24	1	6	1
B10245	41	0	4	1
B10247	23	1	3	0
B10247	23	1	0	1
B10250	58	1	99	1
B10253	66	1	24	0
B10258	30	1	69	1
B10260	63	1	0	1
B10277	24	0	8	1
B10284	32	1	11	1
B10299	26	1	88	0
B10303	46	1	18	0
B10313	35	1	4	1
B10315	52	1	2	0
B10315	52	1	3	1
B10360	28	1	5	0
B10394	26	0	0	1
B10398	59	0	0	0
B10398	59	0	2	1

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

Let Us Compute the Central Tendencies - Median

Arrange in ascending order

The find the value in the middle

2
1
6
4
3
0
99
24
69
0
8
11
88
18
4
2
3
5
0
0
2

0
0
0
0
1
2
2
2
3
3
4
4
5
6
8
11
18
24
69
88
99

10 data points

Value in the middle: MEDIAN

Median divides the sorted data in 2 halves. It is the $\frac{n-1}{2} + 1 = 11^{th}$ data point.

Note n=21 in this dataset.

10 data points

Participant ID	Age	Female	No. of Coughs During Exhaled Breath Collection in 30 min	With Mask
B10176	26	0	2	0
B10206	24	1	1	0
B10206	24	1	6	1
B10245	41	0	4	1
B10247	23	1	3	0
B10247	23	1	0	1
B10250	58	1	99	1
B10253	66	1	24	0
B10258	30	1	69	1
B10260	63	1	0	1
B10277	24	0	8	1
B10284	32	1	11	1
B10299	26	1	88	0
B10303	46	1	18	0
B10313	35	1	4	1
B10315	52	1	2	0
B10315	52	1	3	1
B10360	28	1	5	0
B10394	26	0	0	1
B10398	59	0	0	0
B10398	59	0	2	1

Median tells us that 50% of the people coughed less than or equal to 4 times and 50% more than or equal to 4 times.

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

As of April 03, 2021, what are the Mean and Median # of COVID-19 cases in the world?

#	Country, Other	Total Cases	New Cases	Total Deaths	Total Recovered	Total Tests	Tests/ 1M pop	Population
	World	13,08,44,897	41,919	28,51,025	10,53,43,268			
1	USA	3,13,14,625		5,67,610	2,38,25,635	40,86,74,130	12,29,231	33,24,63,237
2	Brazil	1,29,12,379		3,28,366	1,12,76,628	2,86,00,000	1,33,835	21,36,95,850
3	India	1,23,92,260	1,131	1,64,141	1,15,69,241	24,69,59,192	1,77,643	1,39,01,97,629
4	France	47,41,759		96,280	2,97,734	6,54,82,489	10,01,532	6,53,82,351
5	Russia	45,63,056		99,633	41,86,251	12,09,00,000	8,28,186	14,59,81,647
6	UK	43,53,668		1,26,816	38,73,487	12,75,46,869	18,71,439	6,81,54,420
7	Italy	36,29,000		1,10,328	29,53,377	5,05,89,896	8,37,654	6,03,94,741
8	Turkey	34,00,296		31,892	30,59,462	3,90,70,763	4,59,545	8,50,20,622
9	Spain	32,91,394		75,541	30,54,725	4,27,07,830	9,13,176	4,67,68,434
10	Germany	28,72,379		77,421	25,60,400	5,03,99,177	6,00,092	8,39,85,745
105	Namibia	44,581		531	42,121	3,50,285	1,36,014	25,75,352
106	Ivory Coast	44,445		250	41,281	5,28,303	19,665	2,68,65,781
107	Uganda	40,924		335	40,452	9,45,344	20,193	4,68,15,702
108	Botswana	40,878		591	36,958	9,32,633	3,90,691	23,87,139
109	Jamaica	39,967		607	17,861	2,86,470	96,425	29,70,898
110	Senegal	38,889		1,056	37,532	4,41,464	25,857	1,70,72,997
111	Zimbabwe	36,903		1,524	34,708	3,85,336	25,646	1,50,25,148
112	Malawi	33,639		1,119	30,490	2,19,890	11,277	1,94,98,699
113	Sudan	30,111		2,063	24,214	2,32,965	5,222	4,46,13,862
114	Australia	29,341	8	909	26,299	1,58,28,569	6,15,371	2,57,22,001
115	Malta	29,127		395	28,045	8,13,463	18,38,629	4,42,429
212	Anguilla	25			22	14,579	9,65,241	15,104
213	Saint Pierre Miquelon	24			24	5,126	8,87,927	5,773
214	Montserrat	20		1	19	1,313	2,62,915	4,994
215	Solomon Islands	19			18	4,500	6,434	6,99,457
216	Western Sahara	10		1	8			6,08,301
217	*MS Zaandam*	9		2	7			
218	Marshall Islands	4			4			59,490
219	Samoa	3			2			1,99,406
220	Vanuatu	3			1	470	1,504	3,12,505
221	Micronesia	1			1			1,15,929

There are 221 records

Mean is $\frac{13,08,44,897}{221} = 5,92,058.36$

1. Median is the value of the $\frac{221-1}{2} + 1 = 111^{th}$ record, which is the 36,903 cases in Zimbabwe.

Source: https://www.worldometers.info/coronavirus/
Last accessed: April 03, 2021

DSC 7215

What happens to Mean and Median if we remove the extreme cases of USA, Brazil and India?

	Country, Other	Total Cases	New Cases	Total Deaths	Total Recovered	Total Tests	Tests/ 1M pop	Population
1	France	47,41,759		96,280	2,97,734	6,54,82,489	10,01,532	6,53,82,351
2	Russia	45,63,056		99,633	41,86,251	12,09,00,000	8,28,186	14,59,81,647
3	UK	43,53,668		1,26,816	38,73,487	12,75,46,869	18,71,439	6,81,54,420
4	Italy	36,29,000		1,10,328	29,53,377	5,05,89,896	8,37,654	6,03,94,741
5	Turkey	34,00,296		31,892	30,59,462	3,90,70,763	4,59,545	8,50,20,622
6	Spain	32,91,394		75,541	30,54,725	4,27,07,830	9,13,176	4,67,68,434
7	Germany	28,72,379		77,421	25,60,400	5,03,99,177	6,00,092	8,39,85,745
102	Namibia	44,581		531	42,121	3,50,285	1,36,014	25,75,352
103	Ivory Coast	44,445		250	41,281	5,28,303	19,665	2,68,65,781
104	Uganda	40,924		335	40,452	9,45,344	20,193	4,68,15,702
105	Botswana	40,878		591	36,958	9,32,633	3,90,691	23,87,139
106	Jamaica	39,967		607	17,861	2,86,470	96,425	29,70,898
107	Senegal	38,889		1,056	37,532	4,41,464	25,857	1,70,72,997
108	Zimbabwe	36,903		1,524	34,708	3,85,336	25,646	1,50,25,148
109	Malawi	33,639		1,119	30,490	2,19,890	11,277	1,94,98,699
110	Sudan	30,111		2,063	24,214	2,32,965	5,222	4,46,13,862
111	Australia	29,341	8	909	26,299	1,58,28,569	6,15,371	2,57,22,001
112	Malta	29,127		395	28,045	8,13,463	18,38,629	4,42,429
209	Anguilla	25			22	14,579	9,65,241	15,104
210	Saint Pierre Miquelon	24			24	5,126	8,87,927	5,773
211	Montserrat	20		1	19	1,313	2,62,915	4,994
212	Solomon Islands	19			18	4,500	6,434	6,99,457
213	Western Sahara	10		1	8			6,08,301
214	*MS Zaandam*	9		2	7			
215	Marshall Islands	4			4			59,490
216	Samoa	3			2			1,99,406
217	Vanuatu	3			1	470	1,504	3,12,505
218	Micronesia	1			1			1,15,929

There are 218 records

Mean is $\frac{7,42,25,633}{218} = 3,40,484.56$

1. Median is the value of the $\frac{218-1}{2} + 1 = 109.5^{th}$ record, which is the average of the 109th and the 110th rows, i.e., $\frac{33639+30111}{2} = 31,875$, a value between Malawi's and Sudan's cases.

Source: https://www.worldometers.info/coronavirus/
Last accessed: April 03, 2021

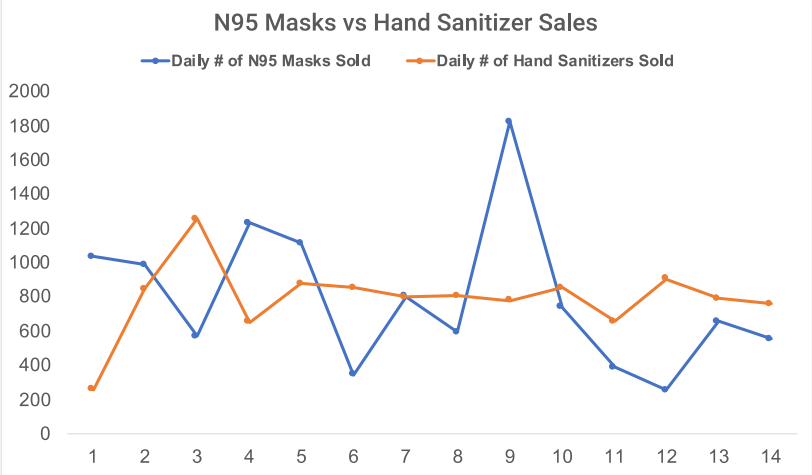
DSC 7215

Central Tendencies (Averages) – Mean and Median

- The **mean** number of cases is 5,92,058 but the **median** is 36,903.
- As the mean is much larger than the median, it indicates that there must be some very large values that are impacting the mean, whereas median, being the middle value, doesn't get impacted.

  Remember, Mean uses all the values (sum) for computation whereas Median is simply the value of the data point in the middle of the sorted dataset.

  - On removing the top 3 countries, the mean dropped dramatically to 3,40,485 but the median shifted slightly to 31,875.
- Such large extreme values are possibly what are called OUTLIERS.



Can you guess which ones are **outliers** in N95 Mask and Hand Sanitizer sales?

DSC 7215

Median

Data #1	Data #2	Data #3
0	0	0
0	0	0
0	0	0
0	0	0
1	1	1
2	2	2
2	2	2
2	2	2
3	3	3
3	3	3
4	4	4
4	4	104
5	5	105
6	6	246
8	8	318
11	8	318
18	8	328
24	8	408
69	9	999
88	9	9999
99	9	99999

Mean
Data #1: 16.6
Data #2: 4.3
Data #3: 5373.4

Note Mean gets pulled in the direction of the outlier.

Same **Median** (4) in all datasets but data is spread and distributed very differently.

We need to understand that too.

Participant ID	Age	Female	No. of Coughs During Exhaled Breath Collection in 30 min	With Mask
B10176	26	0	2	0
B10206	24	1	1	0
B10206	24	1	6	1
B10245	41	0	4	1
B10247	23	1	3	0
B10247	23	1	0	1
B10250	58	1	99	1
B10253	66	1	24	0
B10258	30	1	69	1
B10260	63	1	0	1
B10277	24	0	8	1
B10284	32	1	11	1
B10299	26	1	88	0
B10303	46	1	18	0
B10313	35	1	4	1
B10315	52	1	2	0
B10315	52	1	3	1
B10360	28	1	5	0
B10394	26	0	0	1
B10398	59	0	0	0
B10398	59	0	2	1

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

The Median Can Tell a Lot

Ashes Series 2017-18			
Match	Innings	Sir Alastair Cook	Joe Root
1	1	2	15
1	2	7	51
2	1	37	9
2	2	16	67
3	1	7	20
3	2	14	14
4	1	244	61
5	1	39	83
5	2	10	58
	TOTAL	**376**	**378**
	MEAN	**47**	**42**
	MEDIAN	**14**	**51**

Cook and Root score about the same number of runs in the series.

Based on the Mean, it appears that both Cook and Root performed similarly in the series, with Cook probably slightly better.

Median tells a very different story here. It shows half the time Cook scored less than 14, whereas Root had 50% scores above 51. He was more consistent. The 244 Not Out by Cook pulled the Mean towards this "outlier".

Good Reads for inculcating Statistical thinking in your daily life

http://www.espn.in/cricket/story/_/id/23163239/median-scores-tell-us-batting-careers

Last accessed: March 13, 2020

The Central Tendencies - Mode

No. of Coughs	0	1	2	3	4	5	6	8	11	18	24	69	88	99
Frequency, f	4	1	3	2	2	1	1	1	1	1	1	1	1	1

What is the Mode – the most frequently occurring data point?

The highest frequency in this data is 4. The Mode, therefore, is the value associated with it, which is 0.

That is, 0 is the number that occurs maximum number of times.

Participant ID	Age	Female	No. of Coughs During Exhaled Breath Collection in 30 min	With Mask
B10176	26	0	2	0
B10206	24	1	1	0
B10206	24	1	6	1
B10245	41	0	4	1
B10247	23	1	3	0
B10247	23	1	0	1
B10250	58	1	99	1
B10253	66	1	24	0
B10258	30	1	69	1
B10260	63	1	0	1
B10277	24	0	8	1
B10284	32	1	11	1
B10299	26	1	88	0
B10303	46	1	18	0
B10313	35	1	4	1
B10315	52	1	2	0
B10315	52	1	3	1
B10360	28	1	5	0
B10394	26	0	0	1
B10398	59	0	0	0
B10398	59	0	2	1

The Central Tendencies

Mean and Median need not be in the dataset, but Mode has to be in it.

Median will be in the dataset if there are odd number of data points.

Mode is also the only average that works with categorical data.

Describing Data Through Statistics

Measures of Variability

Descriptive Statistics

Respiratory Virus Shedding in Exhaled Breath and Efficacy of Face Masks

rhinovirus to have a measured fever. Coronavirus-infected participants coughed the most with an average of 17 (s.d. = 30) coughs during the 30-min exhaled breath collection. The profiles of the participants randomized to with-mask versus without-mask groups were similar

Demographic characteristics

Among 55,924 laboratory confirmed cases reported as of 20 February 2020, the median age is 51 years (range 2 days-100 years old; IQR 39-63 years old) with the majority of cases (77.8%) aged between 30–69 years. Among reported cases, 51.1% are male, 77.0% are from Hubei and 21.6% are farmers or laborers by occupation.

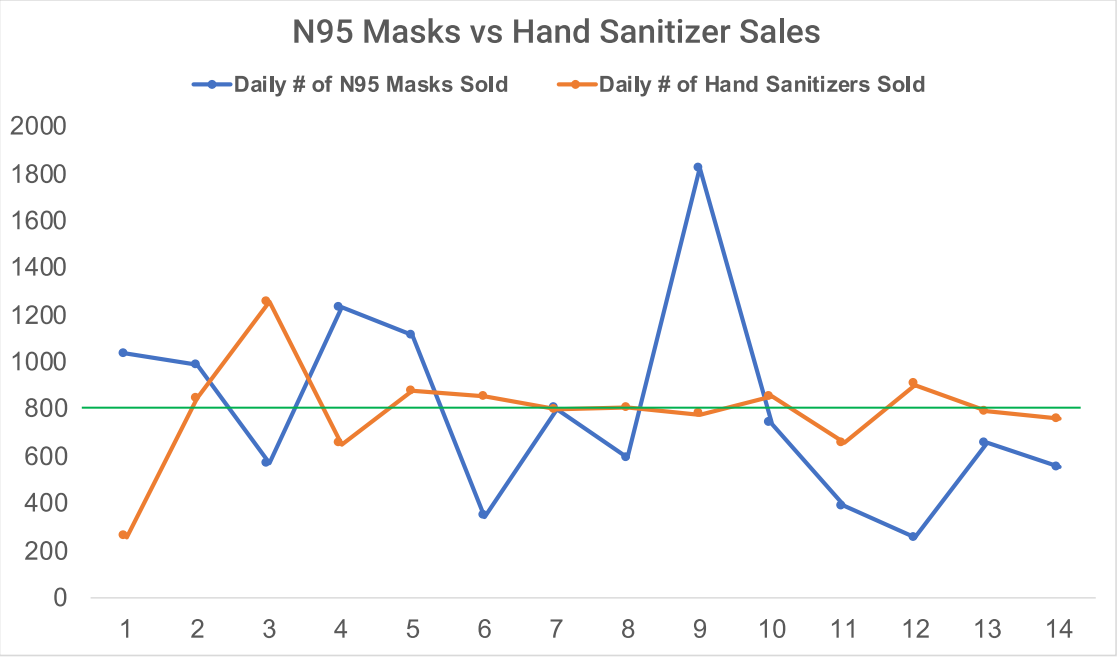
	All who provided exhaled breath	Coronavirus	Influenza virus	Rhinovirus
	(*n* = 246)	(*n* = 17)	(*n* = 43)	(*n* = 54)
	n (%)	*n* (%)	*n* (%)	*n* (%)
Female	144 (59)	13 (76)	22 (51)	30 (56)
Age group, years				
11–17	12 (5)	0 (0)	8 (19)	4 (7)
18–34	114 (46)	10 (59)	11 (26)	24 (44)
35–50	79 (32)	2 (12)	16 (37)	18 (33)
51–64	35 (14)	4 (24)	8 (19)	5 (9)
≥ 65	6 (2)	1 (6)	0 (0)	3 (6)
Chronic medical conditions				
Any	49 (20)	5 (29)	5 (12)	10 (19)
Respiratory	18 (7)	0 (0)	4 (9)	3 (6)
Influenza vaccination				
Ever	94 (38)	6 (35)	15 (35)	20 (37)
Current season	23 (9)	2 (12)	1 (2)	4 (7)
Previous season only	71 (29)	4 (24)	14 (33)	16 (30)
Ever smoker	31 (13)	1 (6)	6 (14)	6 (11)
Time since illness onset, h				
<24	22 (9)	0 (0)	5 (12)	2 (4)
24–48	100 (41)	9 (53)	13 (30)	25 (46)
48–72	85 (35)	8 (47)	18 (42)	20 (37)
72–96	39 (16)	0 (0)	7 (16)	7 (13)
History of measured fever ≥37.8 °C	58 (24)	3 (18)	17 (40)	8 (15)
Measured fever ≥37.8 °C at presentation	36 (15)	2 (12)	18 (42)	2 (4)
Measured body temperature (˚C) at enrollment (mean, s.d.)	36.8 (0.8)	36.9 (0.8)	37.4 (0.9)	36.6 (0.7)
Symptoms at presentation				
Fever	111 (45)	10 (59)	27 (63)	16 (30)
Cough	198 (80)	15 (88)	40 (93)	44 (81)
Sore throat	211 (86)	15 (88)	31 (72)	49 (91)
Runny nose	200 (81)	17 (100)	36 (84)	48 (89)
Headache	186 (76)	13 (76)	30 (70)	38 (70)
Myalgia	176 (72)	12 (71)	31 (72)	34 (63)
Phlegm	176 (72)	9 (53)	34 (79)	41 (76)
Chest tightness	64 (26)	3 (18)	12 (28)	9 (17)
Shortness of breath	103 (42)	6 (35)	14 (33)	25 (46)
Chills	100 (41)	8 (47)	29 (67)	16 (30)
Sweating	95 (39)	5 (29)	18 (42)	20 (37)
Fatigue	218 (89)	16 (94)	38 (88)	48 (89)
Vomiting	19 (8)	2 (12)	5 (12)	2 (4)
Diarrhea	17 (7)	2 (12)	1 (2)	6 (11)
Number of coughs during exhaled breath collection (mean, s.d.)	8 (14)	17 (30)	8 (11)	5 (9)

Sources: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2 and Report of the WHO-China Joint Report on Coronavirus Disease 2019 (COVID-19), 16-24 February, 2020 and https://www.worldometers.info/coronavirus/coronavirus-incubation-period/#asy
Last accessed: April 08, 2020

Descriptive Statistics

Summarize the data by understanding:

1. What is its **average** value – mean, median, mode **(central tendencies)**. These represent the most common value of a variable.
2. How far are data points **spread** – range, IQR (Interquartile Range), variance, standard deviation
3. How are they **distributed** – Box Plot



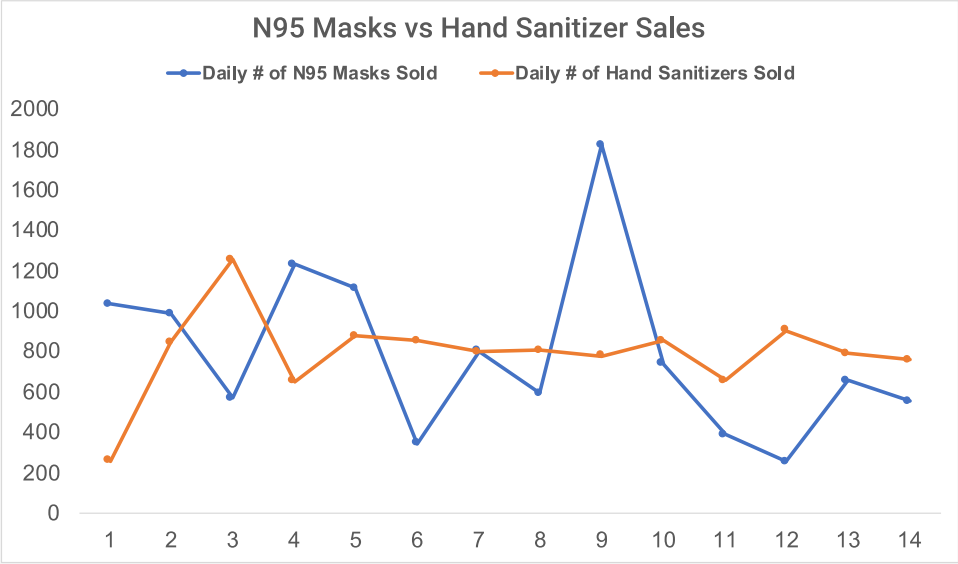
Daily # of N95 Masks Sold	Daily # of Hand Sanitizers Sold	
1034	259	
987	845	
567	1256	
1230	653	
1111	876	
345	854	
802	799	
591	804	
1823	777	
740	853	
389	657	
254	903	
657	789	
555	760	
11085	**11085**	**Sum**
791.8	**791.8**	**Mean**

- The average (mean) number of sales per day is the same for both N95 masks and hand sanitizers.
- However, daily sales of N95 masks fluctuate and are spread a lot more around the mean than the daily sales of hand sanitizers.

DSC 7215

Central Tendencies (Averages) – Mean and Median

- The **mean** number of coughs in a 30-min interval was 17 but the **median** was 4.
- As the mean is much larger than the median, it indicates that there must be some very large values that are impacting the mean, whereas median, being the middle value, doesn't get impacted. *Remember, Mean uses all the values (sum) for computation whereas Median is simply the value of the data point in the middle of the sorted dataset.*
- Such large extreme values are possibly what are called OUTLIERS.



Can you guess which ones are outliers in N95 Mask and Hand Sanitizer sales?

DSC 7215

How to Understand the Spread and Variability

Data #1	Data #2	Data #3
0	0	0
0	0	0
0	0	0
0	0	0
1	1	1
2	2	2
2	2	2
2	2	2
3	3	3
3	3	3
4	4	4
4	4	104
5	5	105
6	6	246
8	8	318
11	8	318
18	8	328
24	8	408
69	9	999
88	9	9999
99	9	99999

Range = Max – Min

Data #1: 99-0 = 99
Data #2: 9-0 = 9
Data #3: 99999-0 = 99999

Participant ID	Age	Female	No. of Coughs During Exhaled Breath Collection in 30 min	With Mask
B10176	26	0	2	0
B10206	24	1	1	0
B10206	24	1	6	1
B10245	41	0	4	1
B10247	23	1	3	0
B10247	23	1	0	1
B10250	58	1	99	1
B10253	66	1	24	0
B10258	30	1	69	1
B10260	63	1	0	1
B10277	24	0	8	1
B10284	32	1	11	1
B10299	26	1	88	0
B10303	46	1	18	0
B10313	35	1	4	1
B10315	52	1	2	0
B10315	52	1	3	1
B10360	28	1	5	0
B10394	26	0	0	1
B10398	59	0	0	0
B10398	59	0	2	1

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

How to Understand the Spread and Variability

Data #1	Data #2	Data #3
0	0	0
0	0	0
0	0	0
0	0	0
1	1	1
2 Q1	2	2
2	2	2
2	2	2
3	3	3
3	3	3
4 Q2	4	4
4	4	104
5	5	105
6	6	246
8	8	318
11 Q3	8	318
18	8	328
24	8	408
69	9	999
88	9	9999
99	9	99999

The IQR (interquartile range) gives the range of the middle 50% of the data.

Just as the median divides the ***sorted*** data into two halves, quartiles divide the ***sorted*** data into 4 quarters. *Likewise, percentiles divide the sorted data into 100 equal parts.*

Quartile 1, Q1: $\left(\frac{1*(n-1)}{4}+1\right) = 6^{th}$ point

Quartile 2, Q2: $\left(\frac{2*(n-1)}{4}+1\right) = 11^{th}$ point (median)

Quartile 3, Q3: $\left(\frac{3*(n-1)}{4}+1\right) = 16^{th}$ point

Note n=21 in this dataset.

IQR = Q3 – Q1 (range for the middle 50% of the sorted data)

Data #1: 11-2 = 9
Data #2: 8-2 = 6
Data #3: 318-2 = 316

Participant ID	Age	Female	No. of Coughs During Exhaled Breath Collection in 30 min	With Mask
B10176	26	0	2	0
B10206	24	1	1	0
B10206	24	1	6	1
B10245	41	0	4	1
B10247	23	1	3	0
B10247	23	1	0	1
B10250	58	1	99	1
B10253	66	1	24	0
B10258	30	1	69	1
B10260	63	1	0	1
B10277	24	0	8	1
B10284	32	1	11	1
B10299	26	1	88	0
B10303	46	1	18	0
B10313	35	1	4	1
B10315	52	1	2	0
B10315	52	1	3	1
B10360	28	1	5	0
B10394	26	0	0	1
B10398	59	0	0	0
B10398	59	0	2	1

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

Median, Range and Interquartile Range

Demographic characteristics

Among 55,924 laboratory confirmed cases reported as of 20 February 2020, the median age is 51 years (range 2 days-100 years old; IQR 39-63 years old) with the majority of cases (77.8%) aged between 30–69 years. Among reported cases, 51.1% are male, 77.0% are from Hubei and 21.6% are farmers or laborers by occupation.

Source: Report of the WHO-China Joint Report on Coronavirus Disease 2019 (COVID-19), 16-24 February, 2020 and https://www.worldometers.info/coronavirus/coronavirus-incubation-period/#asy
Last accessed: April 08, 2020

How to Understand the Spread and Variability

	All who provided exhaled breath	Coronavirus	Influenza virus	Rhinovirus
	(*n* = 246)	(*n* = 17)	(*n* = 43)	(*n* = 54)
Number of coughs during exhaled breath collection (mean, s.d.)	8 (14)	17 (30)	8 (11)	5 (9)

Participant ID	Age	Female	No. of Coughs During Exhaled Breath Collection in 30 min x_i	With Mask
B10176	26	0	2	0
B10206	24	1	1	0
B10206	24	1	6	1
B10245	41	0	4	1
B10247	23	1	3	0
B10247	23	1	0	1
B10250	58	1	99	1
B10253	66	1	24	0
B10258	30	1	69	1
B10260	63	1	0	1
B10277	24	0	8	1
B10284	32	1	11	1
B10299	26	1	88	0
B10303	46	1	18	0
B10313	35	1	4	1
B10315	52	1	2	0
B10315	52	1	3	1
B10360	28	1	5	0
B10394	26	0	0	1
B10398	59	0	0	0
B10398	59	0	2	1

$$\text{Variance} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} = 885.75\left(\frac{coughs}{30\ min}\right)^2$$

$$\text{Standard Deviation, } \sigma = \sqrt{Variance} = 29.8 \cong 30\ \frac{coughs}{30\ min}$$

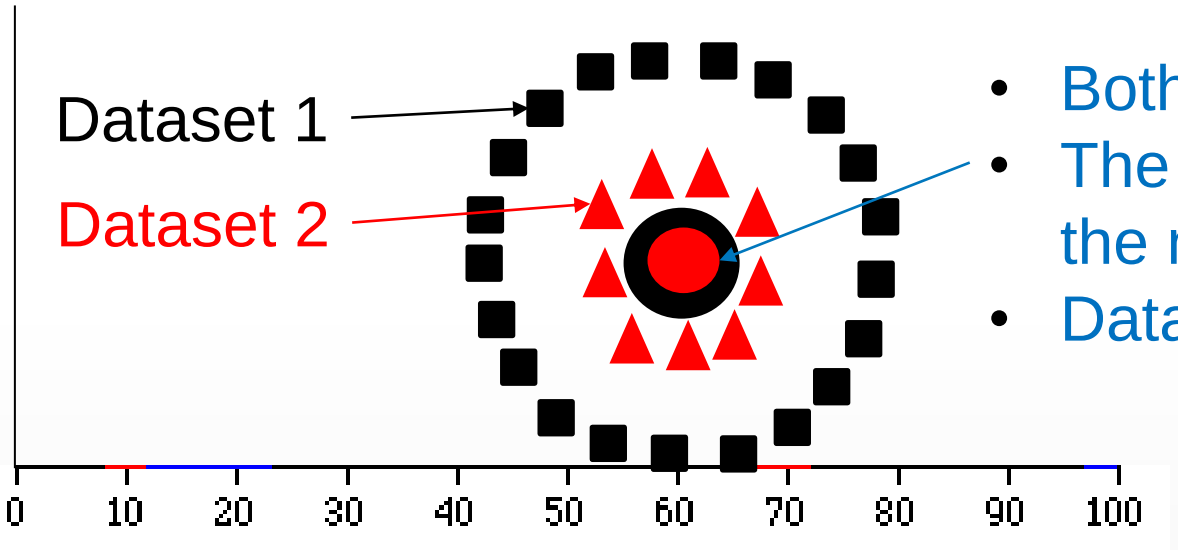
Note n=21 in this dataset (21 records collected from 17 people).

Standard Deviation (or Variance) tells us how far the data points are spread on an average around the mean of the dataset.

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

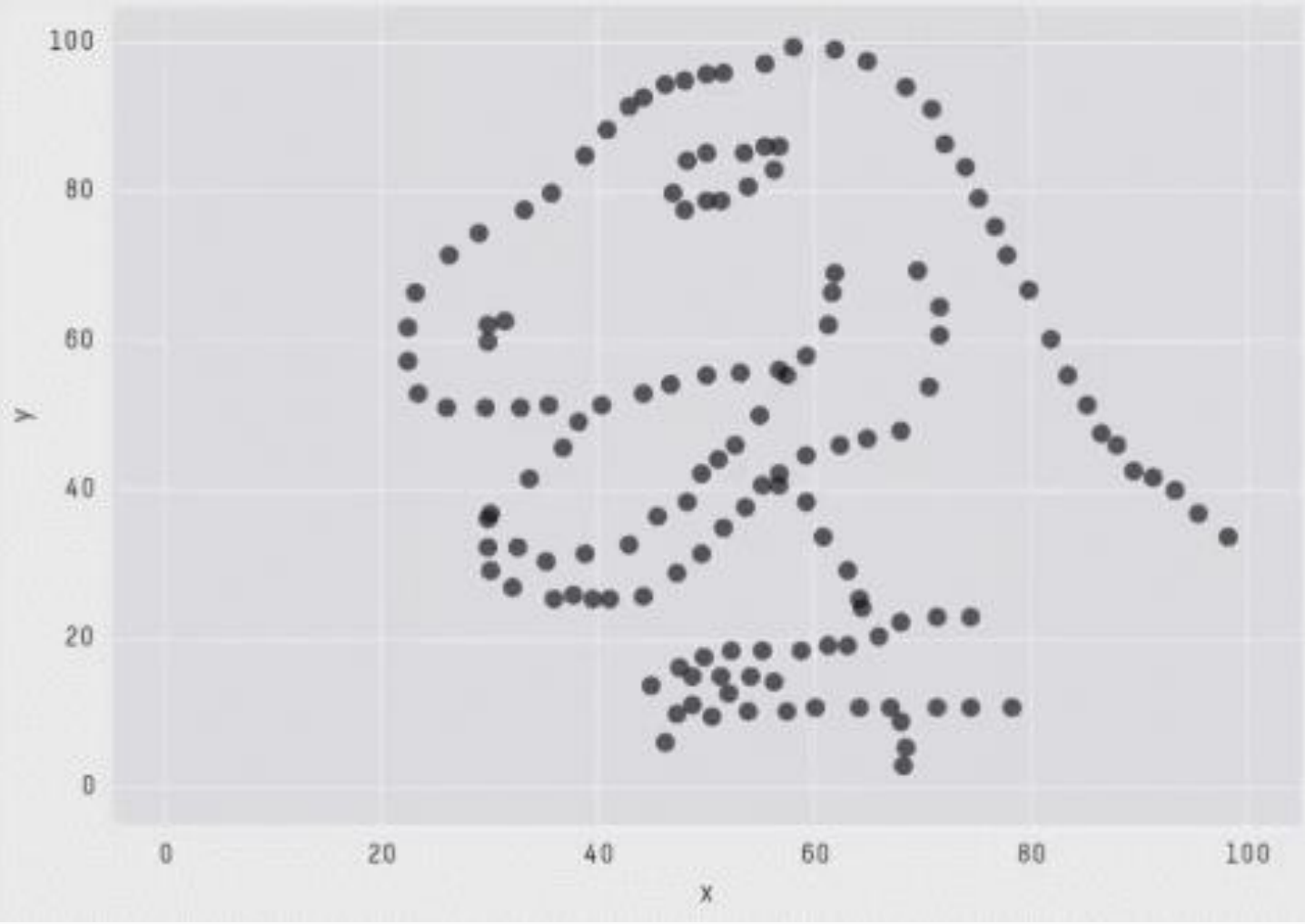
Standard Deviation

Which of the two datasets will have a higher standard deviation?



- Both datasets have same mean.
- The data points are spread closer around the mean in dataset 2 than in dataset 1.
- Dataset 1 has higher standard ***deviation***.

DSC 7215



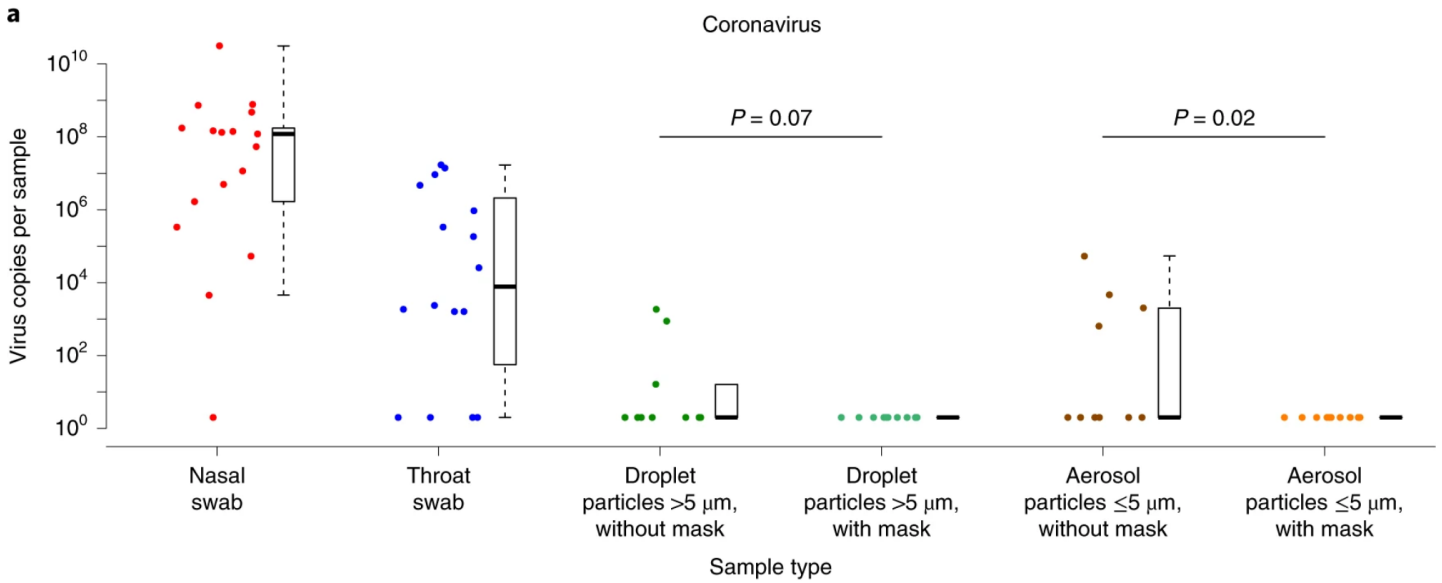
X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Source: https://www.autodeskresearch.com/node/3699
Last accessed: April 09, 2020

Understanding Distribution of the Data

Fig. 1: Efficacy of surgical face masks in reducing respiratory virus shedding in respiratory droplets and aerosols of symptomatic individuals with coronavirus, influenza virus or rhinovirus infection.

From: Respiratory virus shedding in exhaled breath and efficacy of face masks

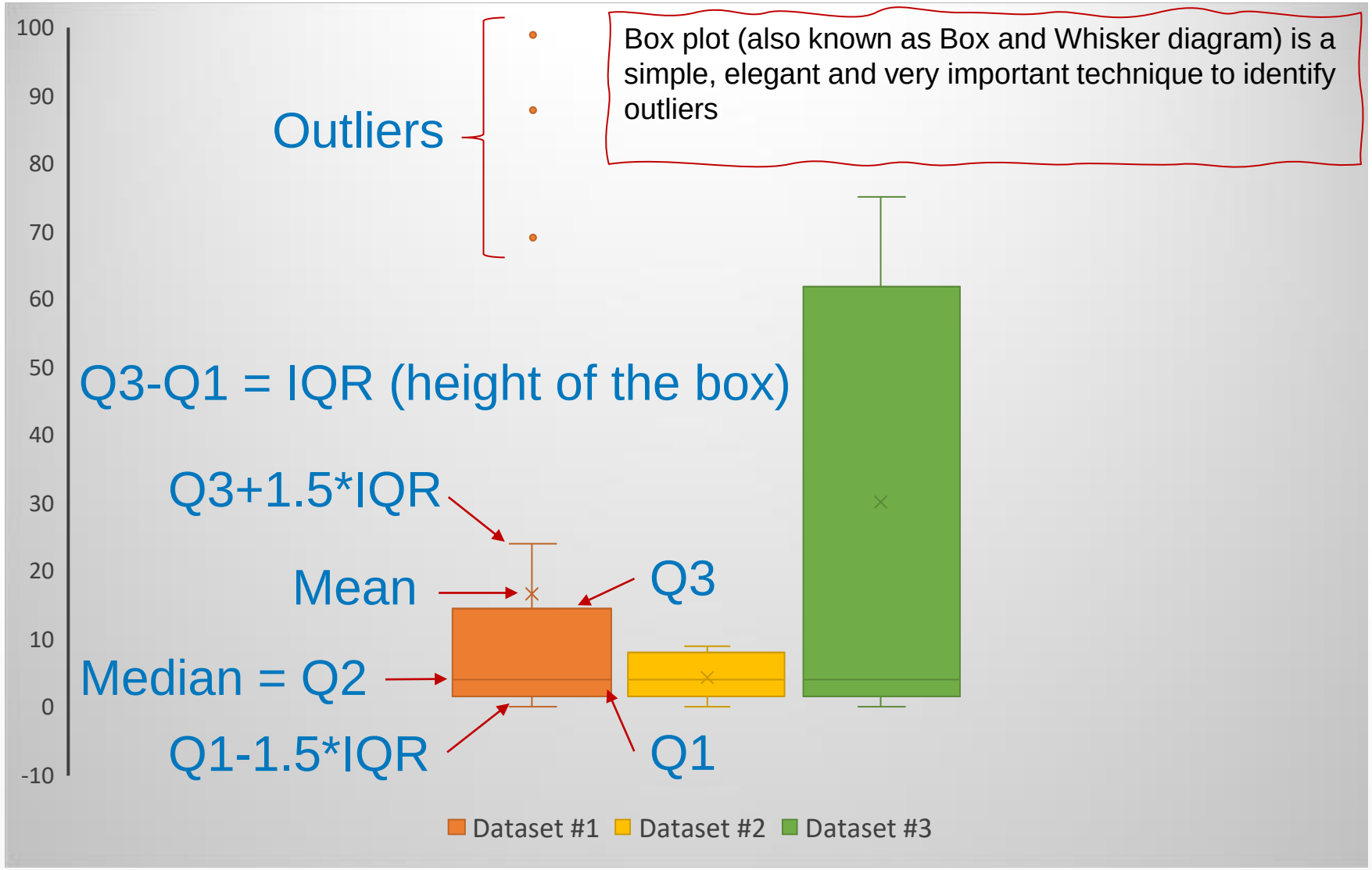


differences in bold. For nasal swabs and throat swabs, all infected individuals were included (coronavirus, n=17; influenza virus, n=43; rhinovirus, n=54). For respiratory droplets and aerosols, numbers of infected individuals who provided exhaled breath samples while not wearing or wearing a surgical face mask, respectively were: coronavirus (n=10 and 11), influenza virus (n=23 and 28) and rhinovirus (n=36 and 32). A subset of participants provided exhaled breath samples for both mask interventions (coronavirus, n=4; influenza virus, n=8; rhinovirus, n=14). The box plots indicate the median with the interquartile range (lower and upper hinge) and ±1.5×interquartile range from the first and third quartile (lower and upper whiskers).

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

Understanding Distribution of the Data – Box Plot

Data #1	Data #2	Data #3
0	0	0
0	0	0
0	0	0
0	0	0
1	1	1
2 Q1	2	2
2	2	2
2	2	2
3	3	3
3	3	3
4 Q2	4	4
4	4	40
5	5	50
6	6	55
8	8	60
11 Q3	8	60
18	8	64
24	8	70
69	9	71
88	9	72
99	9	75



Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

DSC 7215

Understanding Distribution of the Data – Box Plot

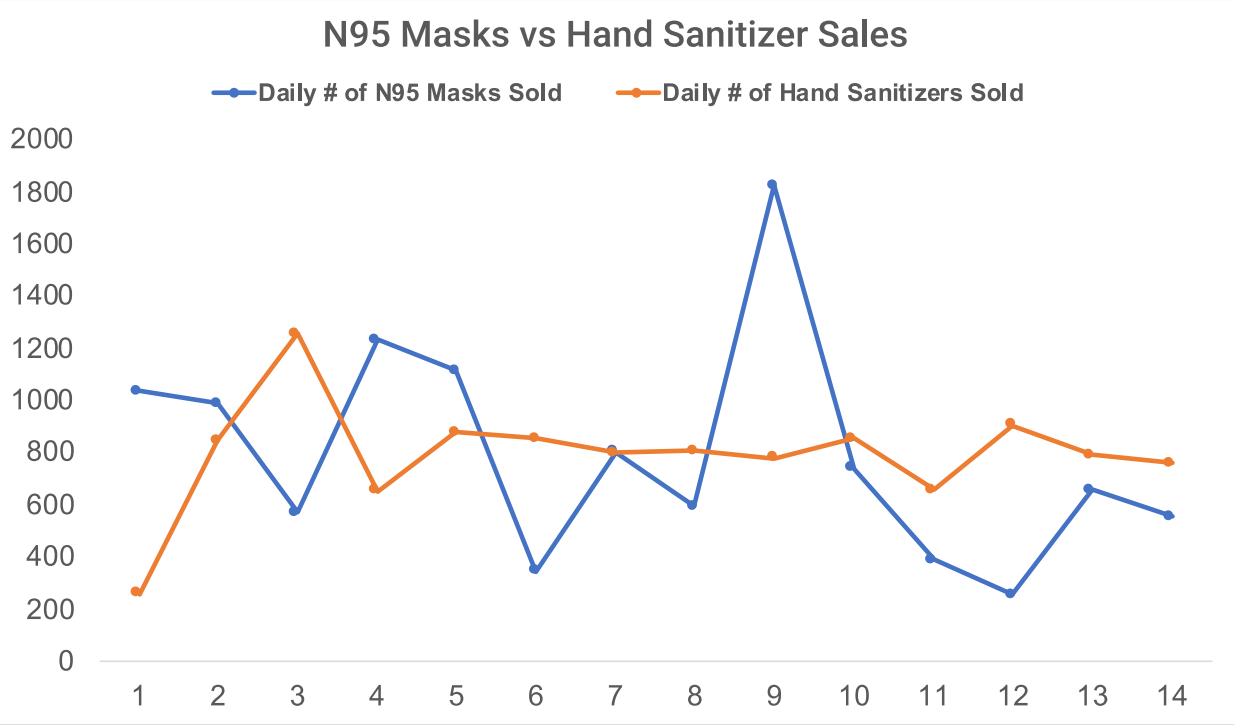
Let us build Box Plots using **Excel** to understand the distribution of the coronavirus viral load collected from nasal and throat swabs

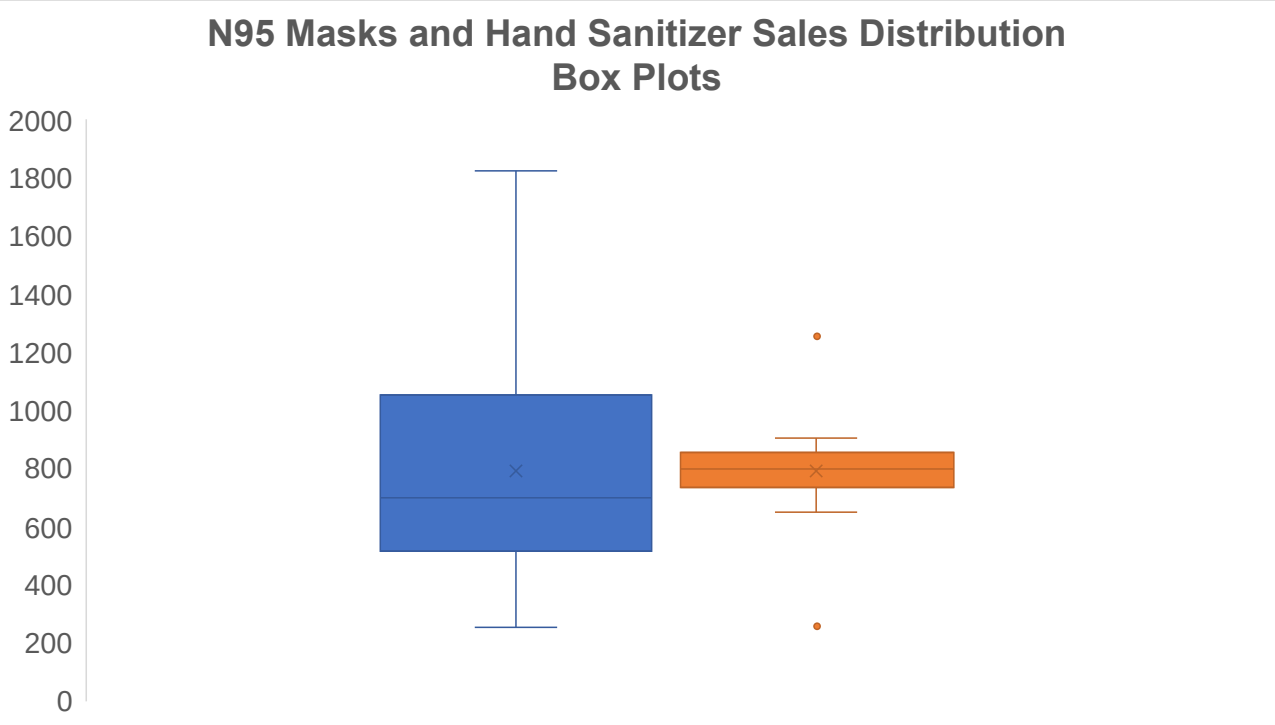
Participant ID	Respiratory Coronavirus Load Identified on **Nasal** Swab (copies per ml)	Respiratory Coronavirus Load Identified on **Throat** Swab (copies per ml)
B10176	292000000.00	1860000.00
B10206	4622481.45	638.73
B10260	12500000000.00	6750000.00
B10277	191000000.00	73500.00
B10284	53200000.00	135000.00
B10299	58970346.88	3674010.31
B10303	47905000.00	10230.00
B10313	56148406.88	637.17
B10315	1990314.65	5518837.50
B10360	69300000.00	748.00
B10394	307050342.50	377729.71

Source: Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0843-2, https://www.nature.com/articles/s41591-020-0843-2
Last accessed: April 08, 2020

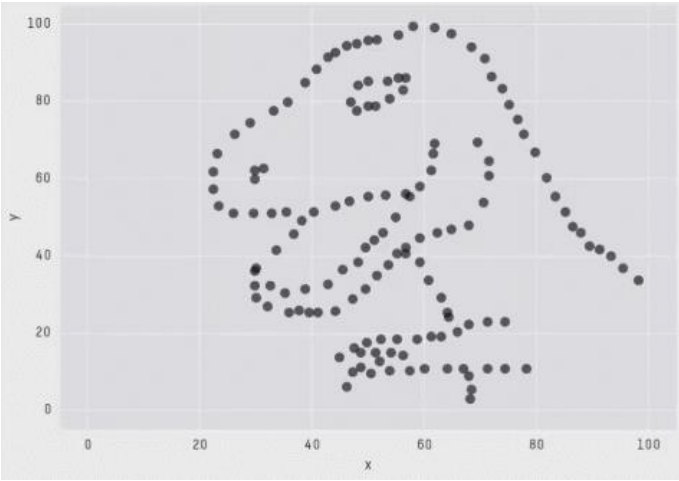
Understanding Distribution of N95 Mask and Hand Sanitizer Sales – Box Plot

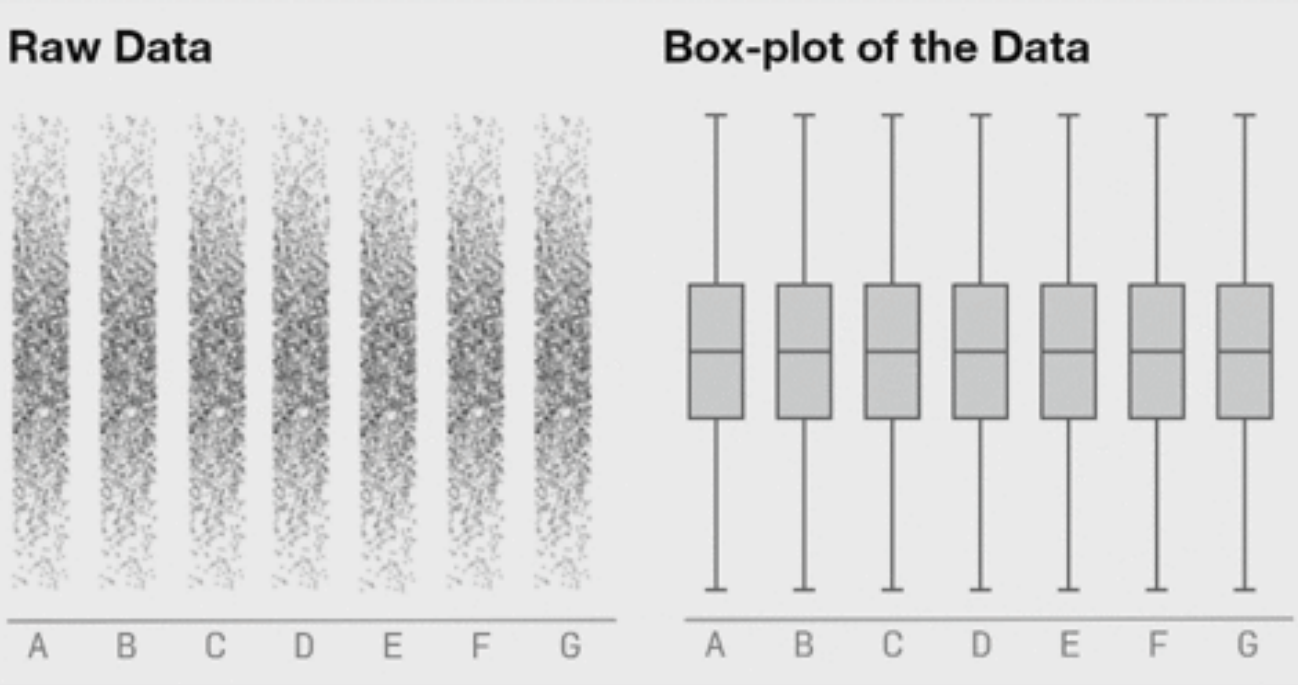
Let us build Box Plots using **Excel** to understand the distribution of the N95 Mask and Hand Sanitizer daily sales



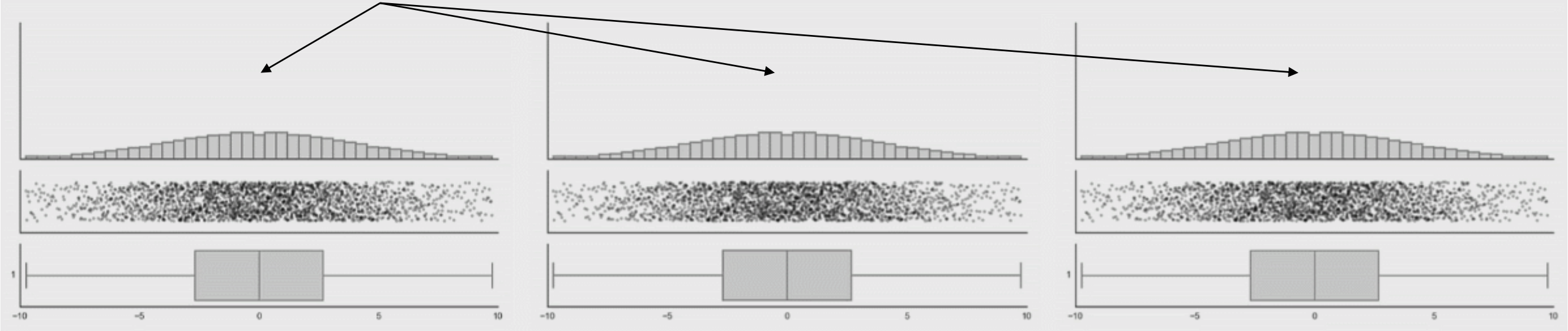


Can you guess which ones are outliers in N95 Mask and Hand Sanitizer sales?



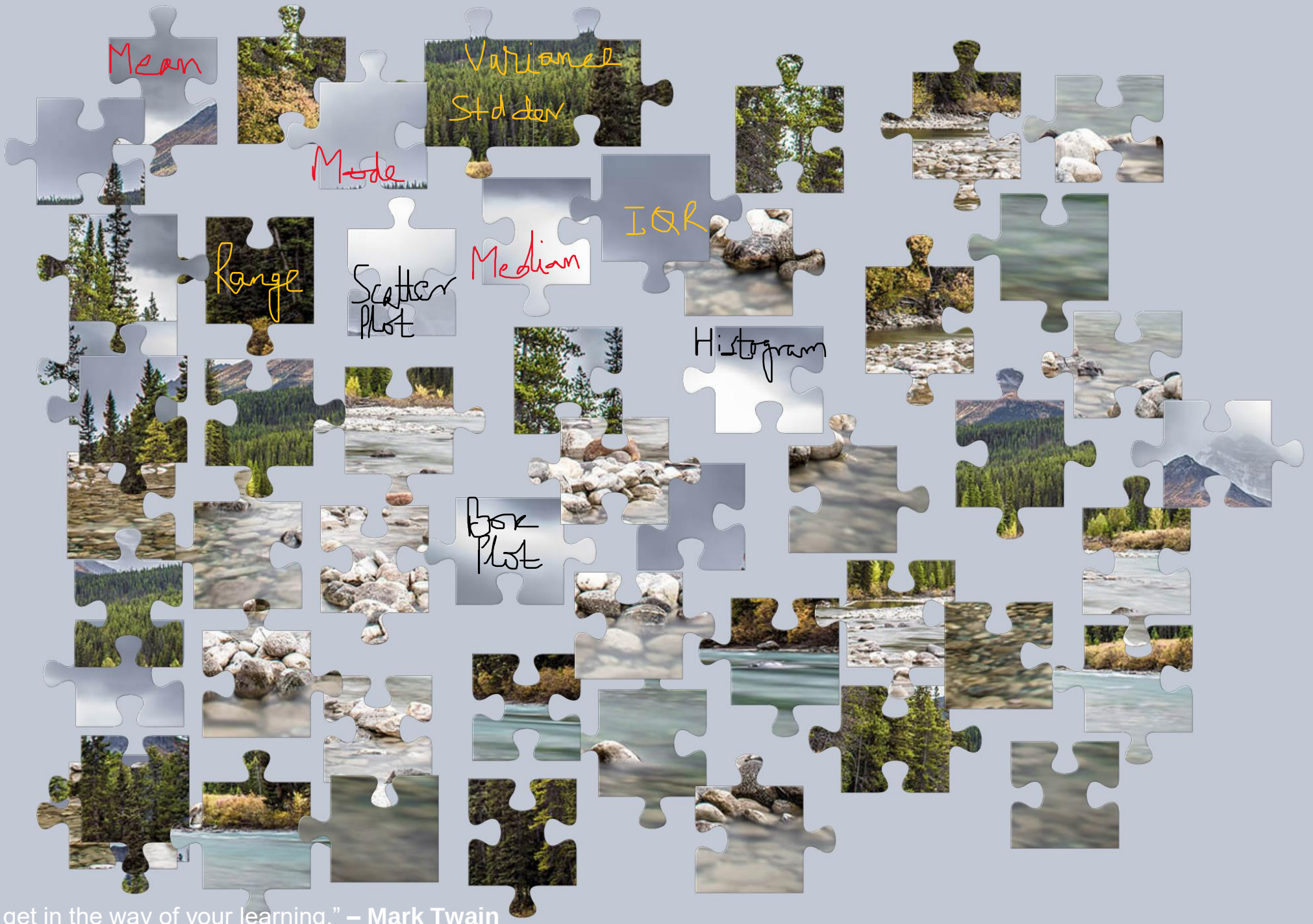


Another way to see data distribution – **Histogram**.
Able to show differences that the Box Plot is unable to.
It has its limitations too.



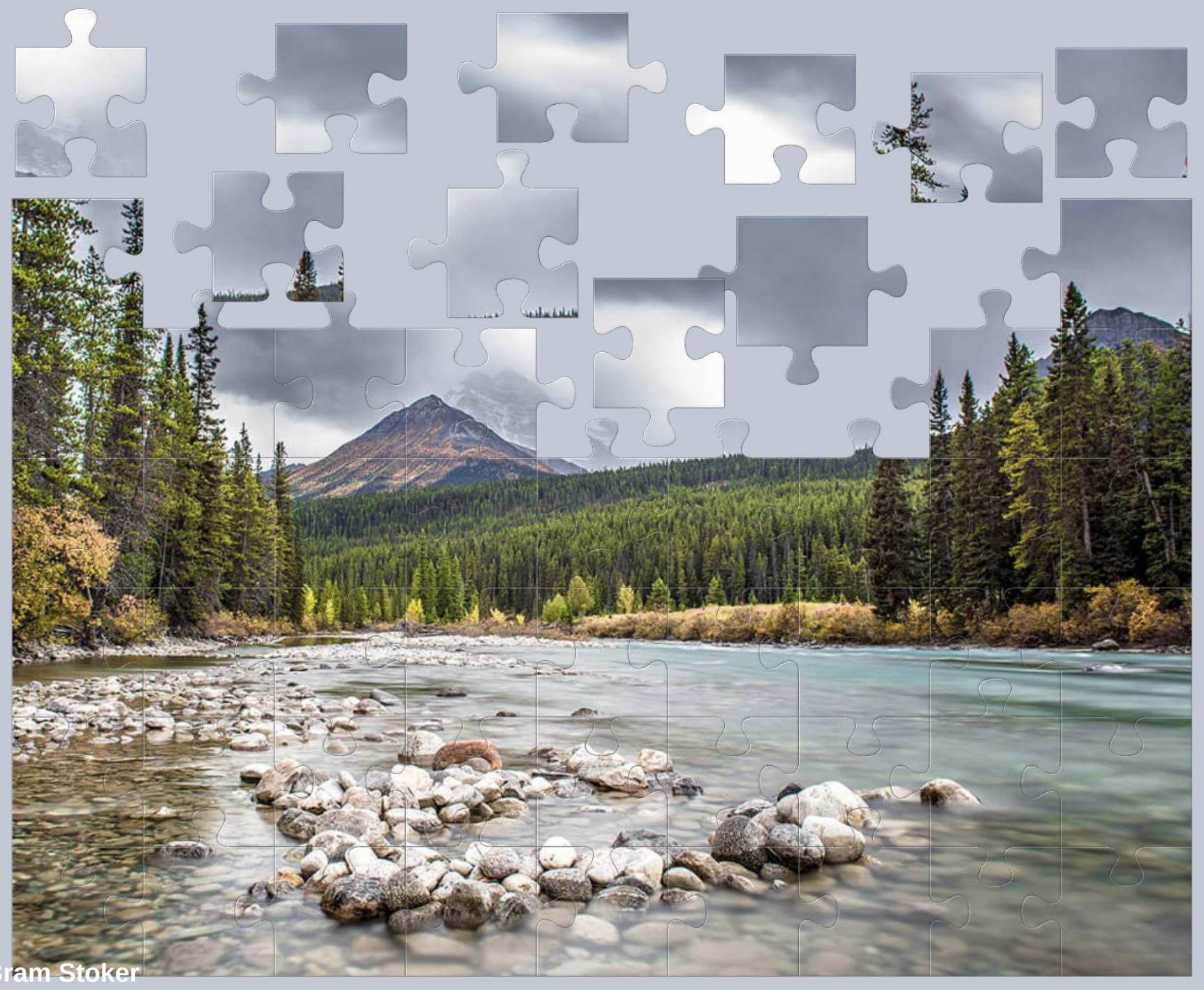
Source: https://www.autodeskresearch.com/node/3699
Last accessed: April 09, 2020

Each Metric is a Piece of the Jigsaw



"Never let formal education get in the way of your learning." **– Mark Twain**

Each Metric is a Piece of the Jigsaw



“We learn from failure, not from success.” **– Bram Stoker**

Each Metric is a Piece of the Jigsaw

“A teacher is one who makes himself progressively unnecessary.” **– Thomas Carruthers**

Learning Outcomes of Today's Session

- These are the expected outcomes from your learning today.
- If you can **confidently** and **fluently** answer the below, you have understood the key aspects.
- Before asking questions, you must first revise the material and the videos of today's session.
- Your questions must clearly indicate **your understanding of the topic** before we clarify.

- LO0: You should be able to appreciate the importance of statistics in daily life pertaining to every domain.
- LO1: You should be able to explain the basic statistical terminology like population, sample, etc., along with some examples.
- LO1.1: You should be able to tell the difference between various data types and give examples of each.
- LO1.2: You should be able to tell the difference between discrete and continuous data and give examples of each.
- LO2: Be able to explain various measures of central tendencies – Mean, Median and Mode.
- LO2.1: Be able to explain when to use median and mode, and the effect of outliers on mean.
- LO3: Be able to define and explain various measures of spread and variability – Range, IQR, Variance and Standard Deviation.
- LO3.1: Be able to construct a Box Plot and identify outliers.

In general, you should be able to appreciate that there are various measures of describing data and no one measure can tell the entire story; you need to compute various statistics to get a better picture.

INSOFE's Vision

The BEST GLOBAL DESTINATION for individuals and organizations to learn and adopt disruptive technologies for solving business and society's challenges.

INSOFE - HYDERABAD
2nd Floor, Jyothi Imperial, Vamsiram Builders
Janardana Hills, Gachibowli
Hyderabad – 500032
+91 93199 77257

INSOFE - BENGALURU
Floors 1-3, L77, 15th Cross Road
Sector 6, HSR Layout
Bengaluru - 560102
+91 93199 77267

INSOFE - MUMBAI
4th Floor - A Wing, Spaces - Kanaki
Andheri-Kurla Road, Chakala
Andheri East, Mumbai - 400093
+91 93199 77269

Email: info@insofe.edu.in **Website:** www.insofe.edu.in

Follow us on Social Media: /insofeglobal/ /school/insofe/ /INSOFEedu /insofe_global/ /InsofeVideos

This presentation may contain references to findings of various reports available in the public domain. INSOFE makes no representation as to their accuracy or that the organization subscribes to those findings.