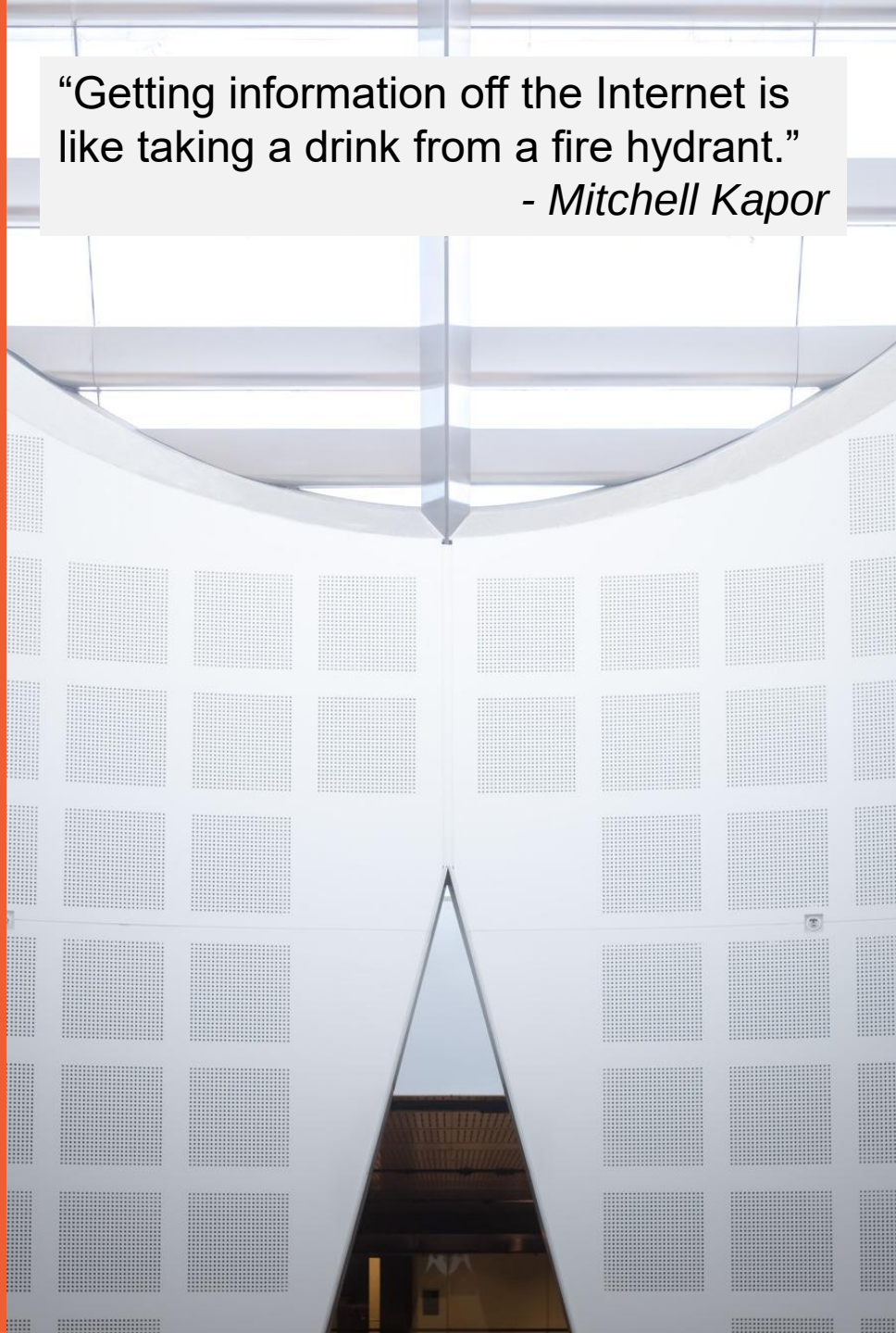
INFO5990: Professional Practice in IT

Week 5:
- Information
- Research
- Estimation

Dr. Mafruha Mowrin Hossain
School of Computer Science

THE UNIVERSITY OF SYDNEY



Quick Overview of Today

Part A: Information

- Finding Information: types; location; purpose
- Trusting Information

Part B: Research

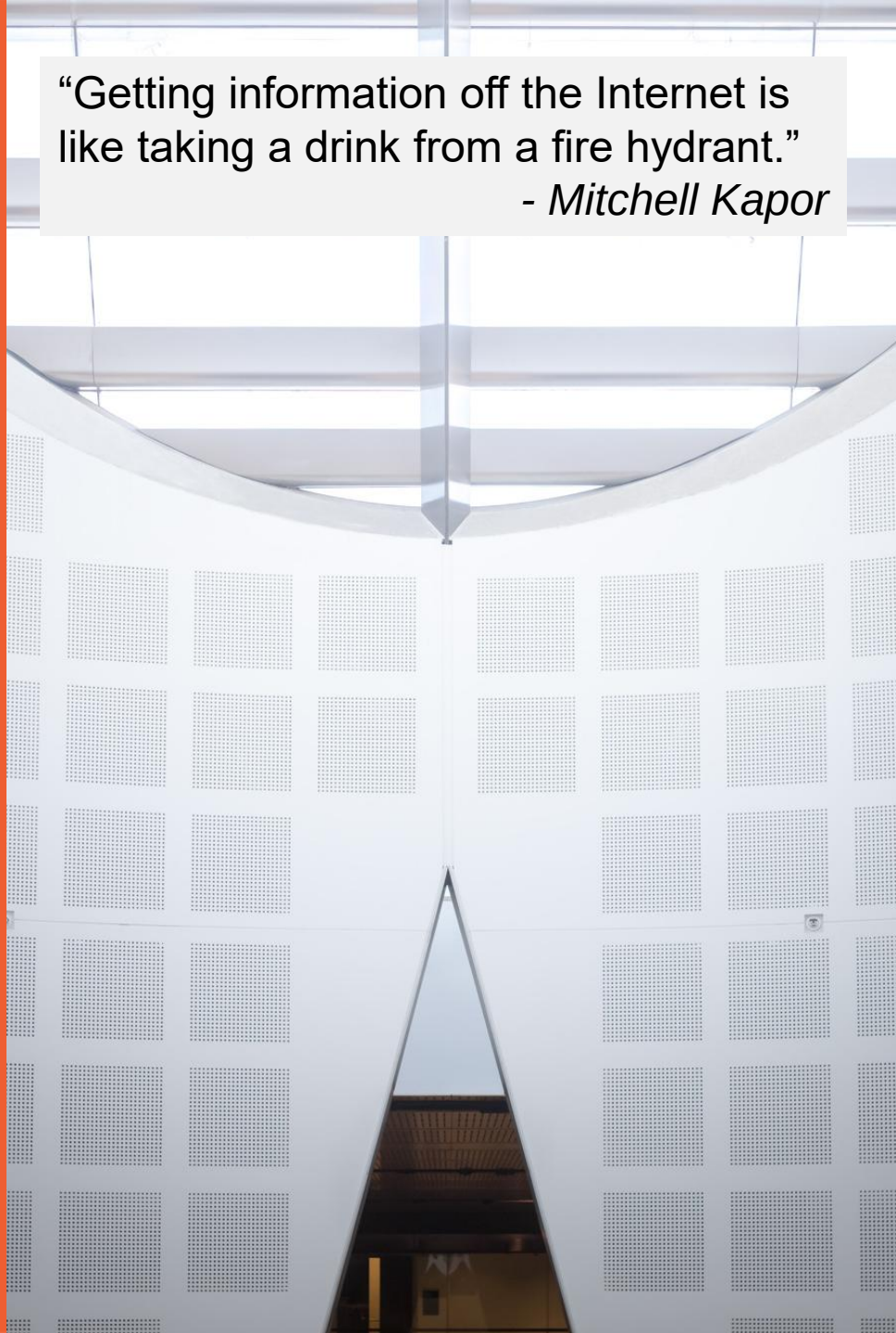
- Types of research
- Business analytics

Part C: Project Estimation

- Understand general approaches to estimating project size ad effort and then explore what this means for organisational IT
- Question: How many of your project failures involved the project taking much more effort / costing much more than expected?

INFO5990: Professional Practice in IT

Week 5:
- Part A: Information



THE UNIVERSITY OF SYDNEY

What is Information

- **What is information?**
  - ✓ Collection of data that can be processed, organised, and structured in a meaningful way to convey knowledge, ideas, or instructions.
- **Can you find?**
  - The name of the Vice-Chancellor of the University of Sydney?
  - His middle name?
  - His wife's name?
  - As the new Vice-Chancellor, what does "success" of his role looks like?

Finding information

- **Sources**
  - The materials from which ideas and information are gathered.
  - Print sources such as books and journals are the most frequently used sources in academic writing.
  - Non-print sources such as music recordings, radio or television broadcasts or transcripts, internet sites, films or images may also be important sources in some disciplines.

- But how do we know what information to trust?


PCMag Australia > Reviews > First Looks > Sales & Marketing > CRM Software

The Best CRM Software for 2021

Cloud CRM can streamline your sales process and grow your customer relationships, and they can do it no matter where folks are located during the pandemic. We test and rank 17 of the top players.

By Molly Mclaughlin, Gadjo Sevilla

6 Aug 2021, 1:19 a.m.

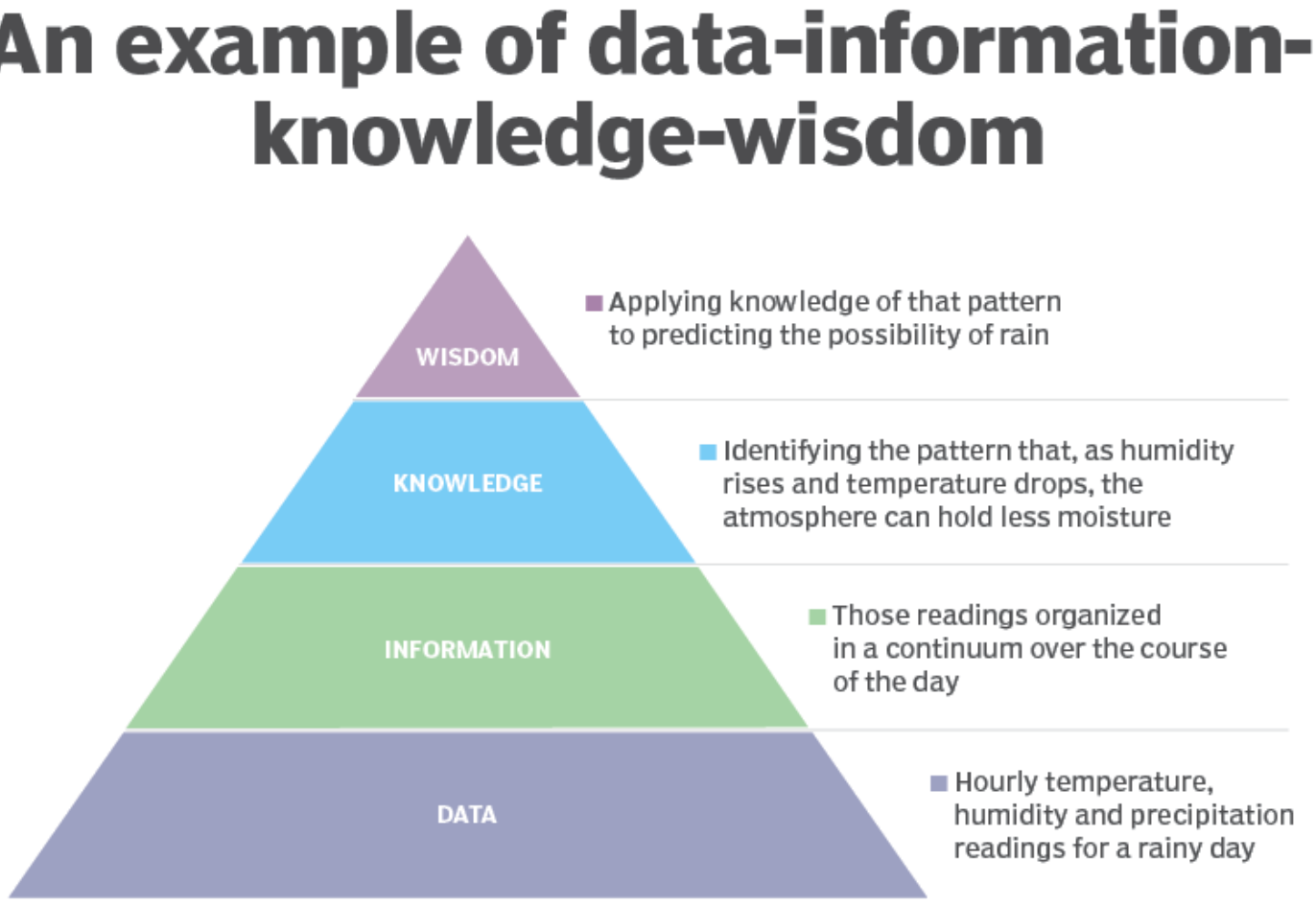
The Best Small Business CRM Software for 2021

The Best Marketing Automation Software for 2021

The Best Email Marketing Software for 2021

TOP PICKS

Data – Information – Knowledge - Wisdom

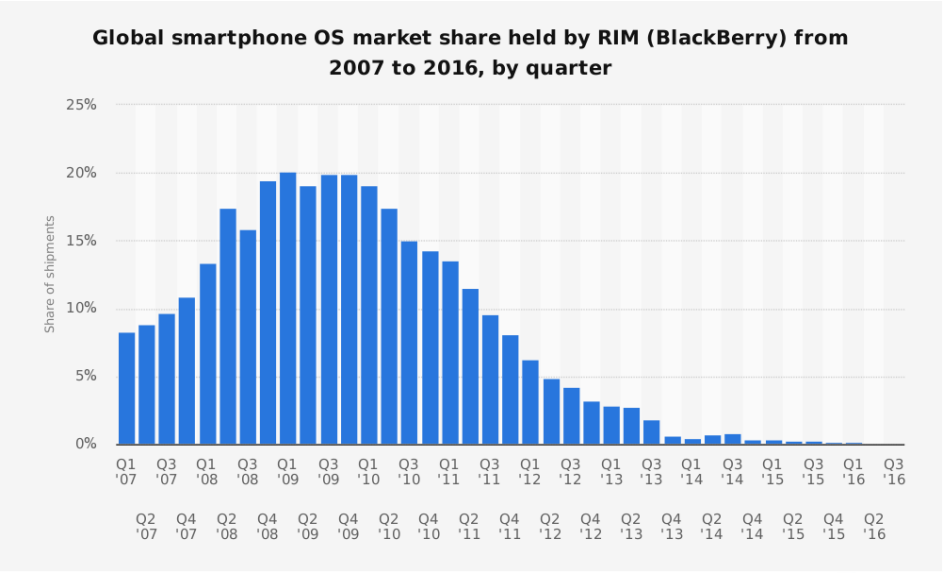


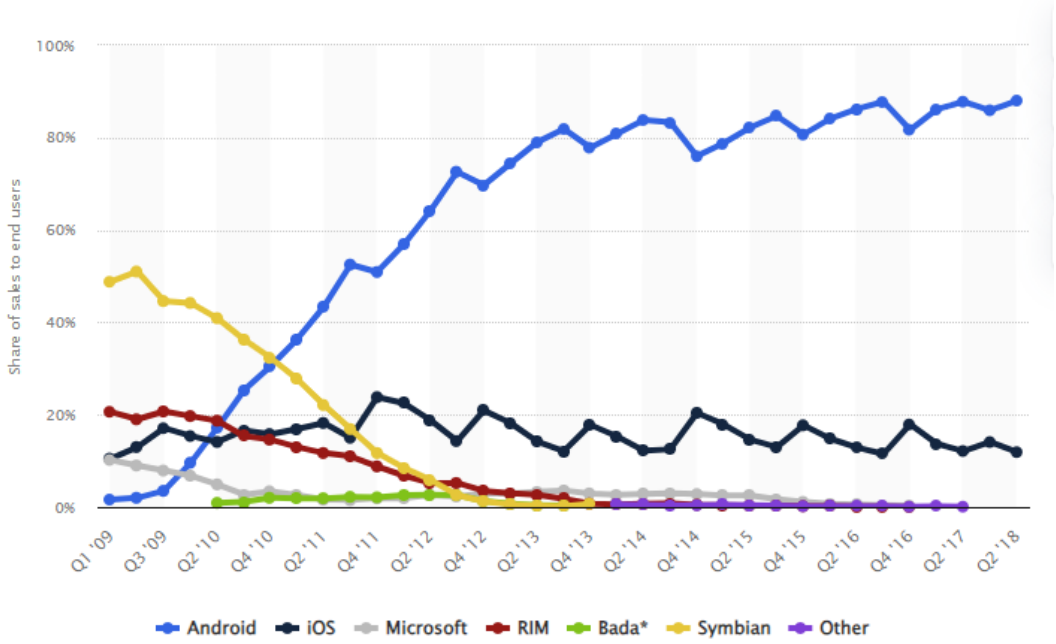
Source: https://www.techtarget.com/searchdatamanagement/definition/information

What level of information / knowledge is needed?

Immediate: narrow technical	• Ability to solve a problem *now* • e.g. *How do I fix a bug in my code?*
Immediate: broader technical	• Ability to manage a project now? • e.g. *How do I integrate Ed into Canvas?*
Possibilities: technical	• Comparing options and capabilities • e.g. *Should we use GSuite or O365?*
Possibilities: business value	• Does something add value? • e.g. *Is it worth investing in a new CRM?*
Future: what changes do I need to be aware of?	• Our field is changing quickly! • e.g. *Will my technical solutions still be relevant in 2 years? 5 years? 10 years?*

Why is this important?





Source: Smartphone OS global market share 2009-2018 | Statista

Sources of information

General

- e.g. https://www.wikipedia.org/

Technical (content)

- e.g. https://www.w3schools.com/

Technical (support)

- e.g. https://stackoverflow.com/

Business

- e.g. https://www.gartner.com/

Research

- E.g. https://dl.acm.org/

What differentiates these?

Information Usefulness

Reliability vs Validity

- Do you know the difference?
  - Information can be reliable (consistency), but not valid(correctness)
  - Information can be unreliable, but still valid
- See https://www.scribbr.com/methodology/reliability-vs-validity/
- (And then explore internal reliability vs external reliability)

Read

- https://www.sagepub.com/sites/default/files/upm-binaries/17810_5052_Pierce_Ch07.pdf
- https://www.academia-research.com/freelance-writing/crediblenon-credible-sources/

Information validity

Different levels

- Primary sources (*My data shows…)*
- Secondary sources (*David claimed that his data showed …)*
- Expert opinion (*Prof Smith thinks that …)*
- Uninformed opinion (*Someone on the bus told me …*)

Evaluating Sources

- Authority of the source
  - Must be reputable and reliable (how do we know)
  - Peer review processes
- Suitability of material
  - Must be related
- Sufficiency of material
  - Include a wide range
  - Both supporting and opposing evidence

What level of information *validity* is needed?

Immediate: narrow technical	• Ability to solve a problem *now* • e.g. *How do I fix a bug in my code?:* • ***Moderate level validity required as It must solve the problem, but doesn't need long-term accuracy***
Immediate: broader technical	• Ability to manage a project now? • e.g. *How do I integrate Ed into Canvas?* • ***High level validity required*** Needs accurate and current steps from reliable sources (e.g., official documentation).
Possibilities: technical	• Comparing options and capabilities • e.g. *Should we use GSuite or O365?* • ***High level validity required*** Requires valid comparisons, specs, and informed trade-offs.
Possibilities: business value	• Does something add value? • e.g. *Is it worth investing in a new CRM?* • ***Very high level validity required*** Must be based on valid business analysis,
Future: what changes do I need to be aware of?	• Our field is changing quickly! • e.g. *Will my technical solutions still be relevant in 2 years? 5 years? 10 years?* • ***Extremely high level validity required*** Needs reliable forecasting, expert insight, and trend data to support long-term planning.

Finding information: Referencing

Purpose:

- Referencing is an essential part of academic writing.
- Its an ethical practice that fulfils the standards of academic conduct that members of a research or scholarly community are expected to uphold.
- Its purpose is to acknowledge the original source of ideas and work that is not the author's own, to point the reader to the original documents so that they can determine independently whether the attributed sources support the author's argument as written, and to help identify the author's own ideas and arguments from that of their sources. http://libguides.library.usyd.edu.au/citation

Different styles:

- See https://libguides.library.usyd.edu.au/citation

APA

- In-text reference (citation):
  - It has been argued that many patients have impaired decision-making when it comes to handling their insomnia (Cheung et al, 2018).
- **Reference:**
- Cheung, J. M. Y., Bartlett, D. J., Armour, C. L., Laba, T. L., & Saini, B. (2016). To Drug or Not to Drug: A Qualitative Study of Patients' Decision-Making Processes for Managing Insomnia. Behavioral Sleep Medicine, 16(1), 1–26. https://doi.org/10.1080/15402002.2016.1163702

Referencing sources of information

Referencing generally has two key elements;

1. An in-text reference that indicates to the reader that a particular concept, phrase, or idea is attributable to someone else,
   - The in-text marker (citation) must be inserted in every sentence that contains an idea or words that come from the original source.
   - Even if you have read something in a source and rephrased it in your own words, you must still cite that source in every sentence in which you have used information from it.
   - For a guide to referencing in-text, see the example at https://canvas.sydney.edu.au/courses/59195/pages/3-referencing-sources-of-information?module_item_id=2332248
2. A complete reference list (in alphabetical order)giving the full details for all sources referred to in the document.

*In this unit you will use the University of Sydney's guidelines for **APA 7th** referencing. Follow this https://canvas.sydney.edu.au/courses/59195/pages/3-referencing-sources-of-information?module_item_id=2332248 for the common source types that you are most likely to use, and for links to further guidelines.*

Plagiarism

- **Consider the following:**

  "Researchers and scholars utilize a wide variety of methods to collect and analyze information. Scientists, for instance, do research by conducting experiments that will support or contradict a theory. Sociologists, on the other hand, use surveys and interviews to gather information from people and draw conclusions on society and culture. Historians study archived texts and artifacts from the relevant time period and make interpretations of the evidence they collect".

  Source: https://www.guide2research.com/research/primary-research-vs-secondary-research

- Would it be OK for you to write:

  Data is important to us when we are trying to make decisions. There are many different ways to obtain the data that we need. For example, we might collect data by conducting experiments, or we might use surveys and interviews to gather information from people, or we could study archived texts and artifacts.

Few words were changed, but it is still **plagiarism** because the structure and key ideas are identical. This is known as "**patchwriting**", where minor edits are made but the original text remains largely unchanged.

Plagiarism

- Consider the following:

> "Researchers and scholars utilize a wide variety of methods to collect and analyze information. Scientists, for instance, do research by conducting experiments that will support or contradict a theory. Sociologists, on the other hand, use surveys and interviews to gather information from people and draw conclusions on society and culture. Historians study archived texts and artifacts from the relevant time period and make interpretations of the evidence they collect".
>
> From:
> https://www.guide2research.com/research/primary-research-vs-secondary-research

- Would it be OK for you to write:

Data is important to us when we are trying to make decision. There are many different ways to obtain the data that we need. For example, we might collect data by conducting experiments, or we might use surveys and interviews to gather information from people, or we could study archived texts and artifacts (Bouchrika, 2021).

References

Bouchrika, Imed (2021), "Primary Research vs Secondary Research: Definitions, Differences, and Examples", accessed at https://www.guide2research.com/research/primary-research-vs-secondary-research, 20th Aug 2021.

The student has added a citation at the end: (Bouchrika, 2021). But they **still copied** the exact wording from the source. Citation alone doesn't make copying okay. You must either:

- **Paraphrase properly** (rewrite in your own words and sentence structure).
- **Use direct quotes** (with quotation marks) if keeping the original words.

Plagiarism

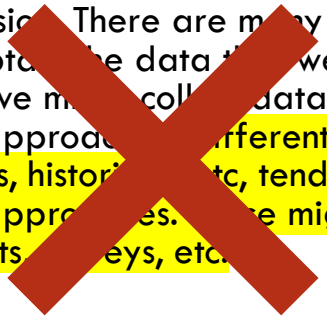
- Consider the following:

  "Researchers and scholars utilize a wide variety of methods to collect and analyze information. Scientists, for instance, do research by conducting experiments that will support or contradict a theory. Sociologists, on the other hand, use surveys and interviews to gather information from people and draw conclusions on society and culture. Historians study archived texts and artifacts from the relevant time period and make interpretations of the evidence they collect".

  From:
  https://www.guide2research.com/research/primary-research-vs-secondary-research

- Would it be OK for you to write:

  Data is important to us when we are trying to make decisi[illegible] There are m[illegible]y different ways to obta[illegible]e data t[illegible]we need. For example, we m[illegible] coll[illegible]data using different approa[illegible]fferent people, such as scientists, histori[illegible]tc, tend to use different appro[illegible]es. [illegible]e might include experiments[illegible]eys, etc.

  The student has tried to change some words but has kept the same structure and meaning as the original.

  This is still plagiarism because:

  1. The wording is too close to the original.
  2. The sentence structure remains the same.
  3. No proper citation is given.

Plagiarism

- Consider the following:

> "Researchers and scholars utilize a wide variety of methods to collect and analyze information. Scientists, for instance, do research by conducting experiments that will support or contradict a theory. Sociologists, on the other hand, use surveys and interviews to gather information from people and draw conclusions on society and culture. Historians study archived texts and artifacts from the relevant time period and make interpretations of the evidence they collect".

From:
https://www.guide2research.com/research/primary-research-vs-secondary-research

- Would it be OK for you to write:

Data is important to us when we are trying to make decision. There are many different ways to obtain the data that we need. For example, we might collect data using different approaches. Different people, such as scientists, historians, etc, tend to use different approaches. These might include experiments, surveys, etc. (Bouchrika, 2021).

References

Bouchrika, Imed (2021), "Primary Research vs Secondary Research: Definitions, Differences, and Examples", accessed at https://www.guide2research.com/research/primary-research-vs-secondary-research, 20th Aug 2021.

- The student has completely changed the sentence structure.
- The key ideas are still there, but expressed in their own words.
- A proper citation is included (Bouchrika, 2021).

Plagiarism

- Consider the following:

> "Researchers and scholars utilize a wide variety of methods to collect and analyze information. Scientists, for instance, do research by conducting experiments that will support or contradict a theory. Sociologists, on the other hand, use surveys and interviews to gather information from people and draw conclusions on society and culture. Historians study archived texts and artifacts from the relevant time period and make interpretations of the evidence they collect".
>
> From:
> https://www.guide2research.com/research/primary-research-vs-secondary-research

- Would it be OK for you to write:

Data is important to us when we are trying to make decision. There are many different ways to obtain the data that we need. For example, we might collect data "conducting experiments, [or we might] use surveys and interviews to gather information from people [or we could study archived texts and artifacts" (Bouchrika, 2021).

References

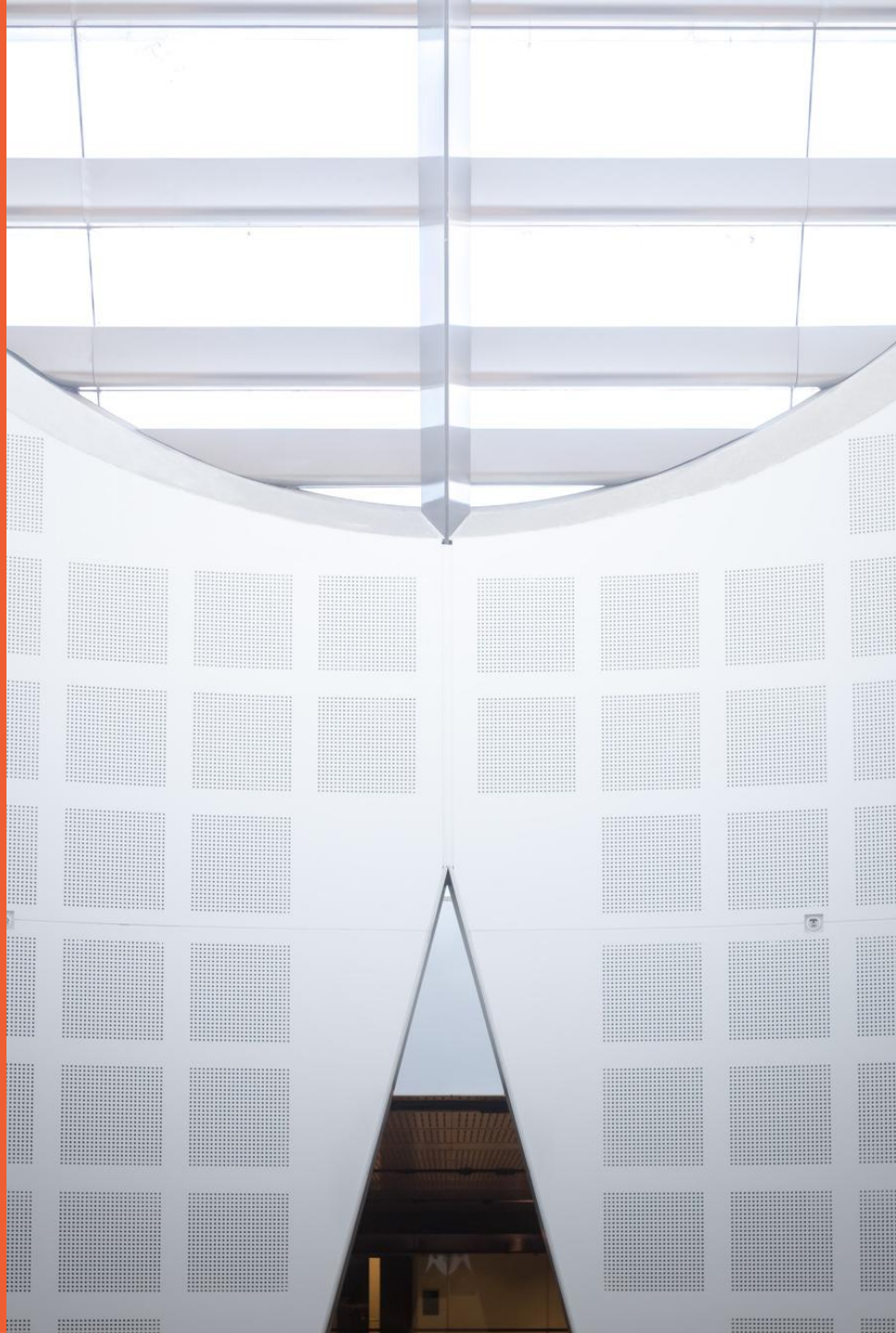
Bouchrika, Imed (2021), "Primary Research vs Secondary Research: Definitions, Differences, and Examples", accessed at https://www.guide2research.com/research/primary-research-vs-secondary-research, 20th Aug 2021.

- The student uses quotation marks for the copied words.
- They include a proper citation (Bouchrika, 2021).
- They still explain the idea in their own words before and after the quote.

INFO5990: Professional Practice in IT

Week 5

Part B: Research



THE UNIVERSITY OF SYDNEY

Research - why

- **Why is it important for IT professionals?**
- **Where would you use research techniques?**

- Help make informed decisions
- Need to produce research in career
- Evaluating research in the media
- Enabling professionals to stay relevant, innovate, and contribute to their organizations' and industry's growth and success.

Primary Research

- Involves collecting original data directly from sources.
- Used when specific, up-to-date, or firsthand information is needed.
- Provides more control over how data is collected and analyzed.
- Often used in academic research, product testing, and customer feedback.
- **Examples: surveys, interviews, observations, experiments.**

Secondary Research

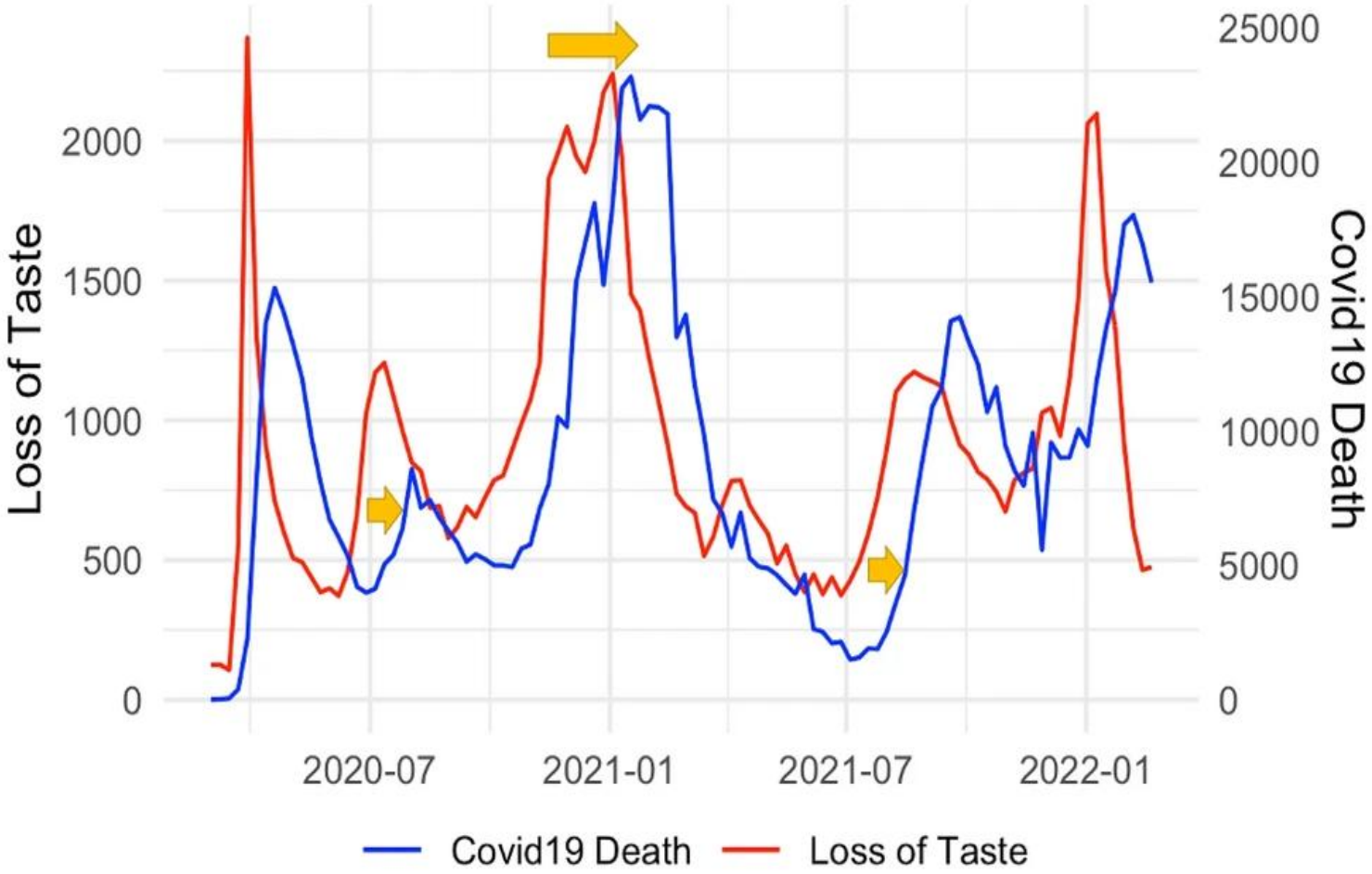
- Involves using existing data collected by others.
- Quicker and less expensive than primary research.
- Useful for getting background information or identifying trends.
- May be outdated or not specific to your exact topic.
- **Examples: books, articles, websites, government reports, databases.**

Why Statistics Are Important for Research

- Helps collect, analyze, and interpret data accurately.
- Supports evidence-based conclusions and decisions.
- Identifies patterns, trends, and relationships in data.
- Validates hypotheses and tests assumptions scientifically.
- Enhances credibility, objectivity, and reliability of research findings.
- Enables prediction and generalization to larger populations.

Knowing about statistics

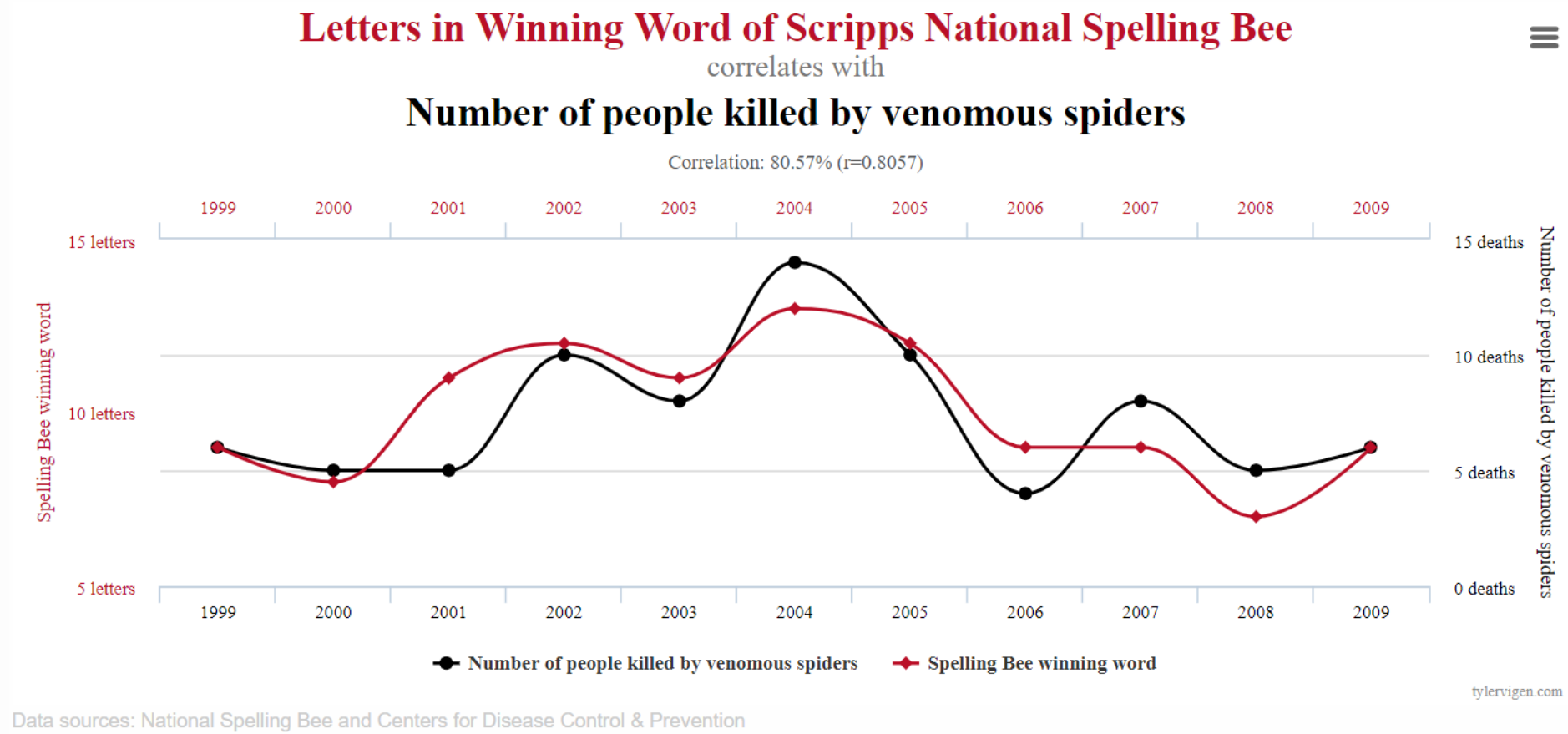
From: COVID-19 forecasts using Internet search information in the United States



- Every time loss of taste **spiked,** COVID-19 deaths increased shortly after.
- This tells us that **search data** (e.g., people Googling "loss of taste") **can predict real-world outbreaks.**

Source: https://www.nature.com/articles/s41598-022-15478-y/figures/1

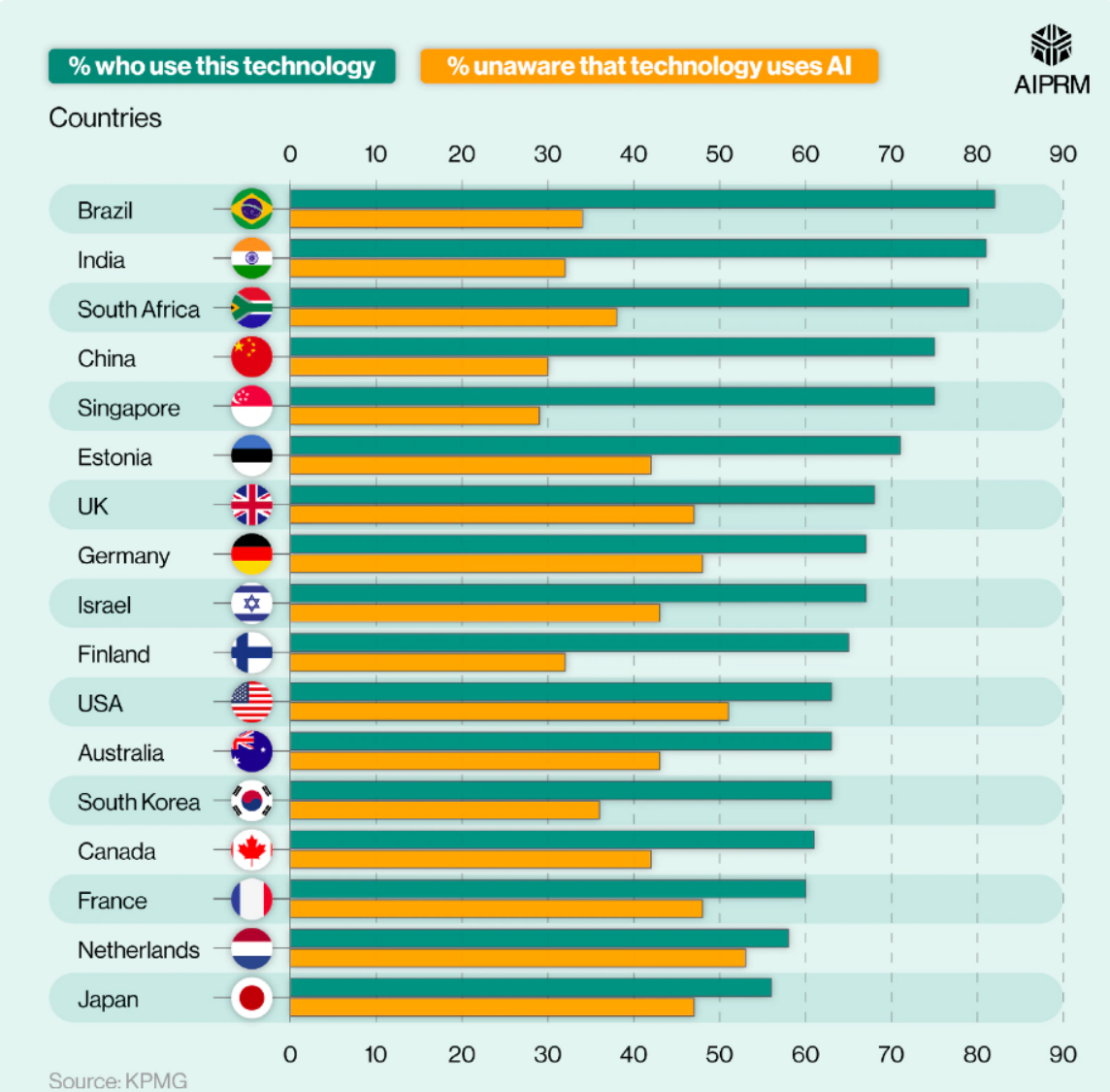
Knowing about statistics



- The number of letters in the winning word of the Scripps National Spelling Bee.
- The number of people killed by venomous spiders.
- **Spelling bees are causing spider attacks?**

Knowing about statistics

Source: https://www.aiprm.com/ai-statistics/



Knowing about statistics

Common mistakes

- **Data ≠ Insight** :Just having numbers doesn't mean they tell a useful story.
- **Correlation ≠ Causation** :not everything that happens together is connected.
- **Poorly framed questions**: Asking biased or vague survey questions can lead to misleading results.
- **Sample sizes and statistical significance**: If only 10 people were surveyed in a study, can we trust its conclusions?

Have a read of:

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6785265/
- **You must sign in LinkedIn to access the following link**
- https://www.linkedin.com/pulse/how-lie-numbers-kevin-gray/

Business Analytics

- **Consider a scenario:**
  - A national fast-food chain (BigBurgers) wants to open three new stores, and is trying to decide the best locations…
  - What data could they collect that might help them understand where to locate the new stores?
  - What evidence could the company collect to help them determine whether the stores have been *successful*?

Business Analytics

Development

- Focuses on the effectiveness of the development methodology
- Metrics: e.g. Project velocity (sprints?)
- https://techbeacon.com/app-dev-testing/9-metrics-can-make-difference-todays-software-development-teams

Technical Operations

- Focuses on the effectiveness of the solution
- Metrics: e.g. Bug reports/tickets

Business Operations

- Focuses on the value provided
- Metrics: e.g. productivity; client/customer satisfaction; …

But remember: data ≠ insight

Top Business Analytics Tools

- Power BI (by Microsoft)
- Tableau
- SAP
- SAS Business Intelligence
- IBM Cognos Analytics etc.

INFO5990: Professional Practice in IT

Week 5:
Part C: Project Estimation

THE UNIVERSITY OF SYDNEY

Can you estimate how much "effort" it would take…

- To write a "Hello World" program in Python?
- To create a simple calculator iPhone app?
- To generate the software for a modern car?
- To write the control software for the Falcon 1 rocket?
- To completely redesign Facebook?
- To create Alibaba?
- To implement a payroll system for a state Health department?

Why is it difficult?

Five steps in project estimation

1. Determine the SIZE of the project
   - software metrics: lines of code, function points
2. Determine the EFFORT required
   - Person hours, days, weeks or months
3. Decide on the RESOURCES needed
   - e.g. how many engineers or programmers
4. Calculate the DURATION
   - e.g. 20 person-hours, 3 people:
     ∴ DURATION = 20 / 3 = 6.3 hours
5. Calculate the COST
   - e.g. 20 person-hours at $70 per hour:
     ∴ COST = $1,400

Six approaches to project estimation

Function point analysis- Estimates based on software **functions**

Algorithmic cost models- Uses **mathematical models**

Component matrix- Estimates based on **modular components**

Expert judgement- Ask experienced developers (widely used!)

Sum of the parts- Breaks project into smaller tasks and sums them up.

Estimation by analogy- Compare with past projects.

1. Function Points Metric

- **Function Points Metric** is used for Measuring Software Size and Estimation.
- It is independent of:
  - Programming language
  - Technology used
  - Development methodology
- It helps in:
  - Estimating development effort
  - Measuring team productivity
  - Comparing software projects

1. Function Points Metric

Components of Function Point Analysis

1. **Unadjusted Function Points (UFP):** Calculated by identifying and counting 5 types of system components:
    - External Inputs (EI)
    - External Outputs (EO)
    - External Inquiries (EQ)
    - Internal Logical Files (ILF)
    - External Interface Files (EIF)
2. **Technical Complexity Factor (TCF):**
    - Adjustment factor based on 14 technical and environmental characteristics
3. **Final Function Points (FP):**
    - Calculated using:

$$\mathbf{FP = UFP \times TCF}$$

1. Function Points Metric

- **Unadjusted Function Points (UFP)**
- For any product, size in "function points":

$$UFP = 4 \times Inp + 5 \times Out + 4 \times Inq + 10 \times Maf + 7 \times Inf$$

Inputs | Outputs | Inquiries | Master Files | Interfaces

Function Type	Description
External Inputs (EI)	User input data
External Outputs (EO)	System-generated reports or outputs
External Inquiries (EQ)	User-initiated input/output queries
Internal Logical Files (ILF)	Data maintained by the system
External Interface Files (EIF)	Data used but not maintained

Function Points Metric

- 1. **Classify each component** (Inp, Out, Inq, Maf, Inf) **as** simple, average, or complex.
  - Assign number of function points.
  - Sum yields the UFP (unadjusted function points).

	Level of Complexity		
Component	**Simple**	**Average**	**Complex**
Input item	3	4	6
Output item	4	5	7
Inquiry	3	4	6
Master file	7	10	15
Interface	5	7	10

Function Points Metric

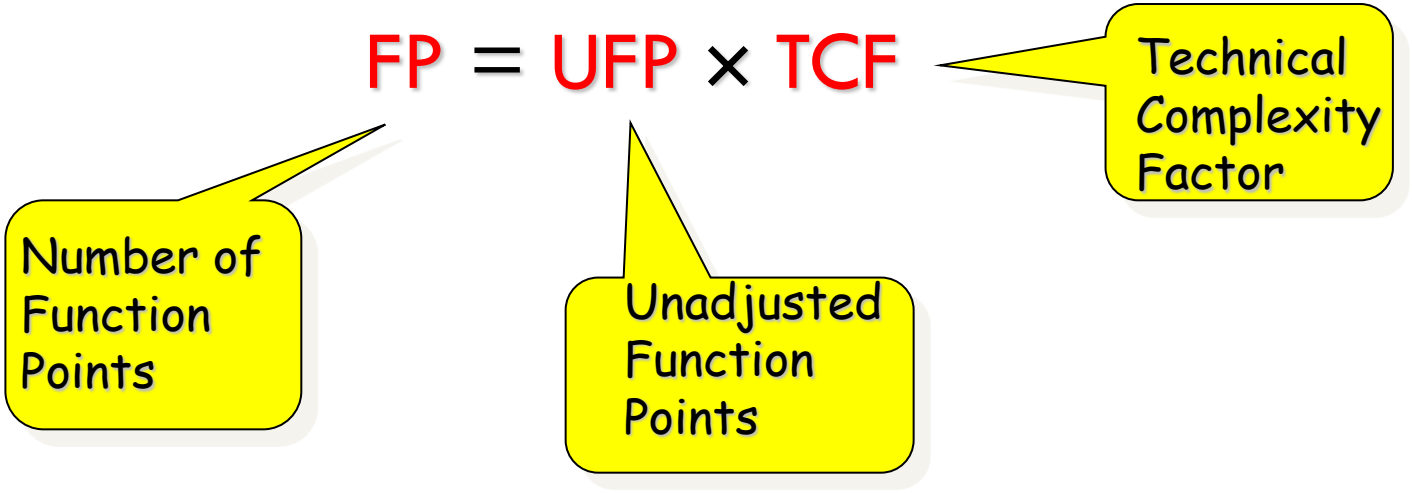
- 2. **Compute technical complexity factor (TCF)**
  - TCF considers 14 factors, each rated from 0 (no influence) to 5 (strong influence):
  - Add 14 numbers = total degree of influence (DI)

    $$TCF = 0.65 + 0.01 \times DI$$
  - TCF in range [0.65…1.35]

1. Data communication
2. Distributed data processing
3. Performance criteria
4. Heavily utilized hardware
5. High transaction rates
6. Online data entry
7. End-user efficiency
8. Online updating
9. Complex computations
10. Reusability
11. Ease of installation
12. Ease of operation
13. Portability
14. Maintainability

Function Points Metric

- **3. Number of function points:**



From: https://www.geeksforgeeks.org/software-engineering-calculation-of-function-point-fp/

Function Points Metric (Example: Hospital Management System)

UFP Calculation

Feature	Type	Description
Patient registration form	Input	Add new patient details
Appointment booking form	Input	Book appointments with doctors
Staff login	Input	Doctors and nurses log in
Medical record entry form	Input	Enter diagnosis and treatment details
View patient profile	Output	Show full patient details
View appointment schedule	Output	Daily/weekly appointment list
Generate discharge summary	Output	Final treatment summary after discharge
Search patients	Inquiry	Find patients by ID or name
Check assigned patients	Inquiry	Show patients under a specific doctor/nurse
Patient database	Master File	Stores patient medical history
Staff database	Master File	Stores doctor and nurse details
External pharmacy system API	Interface	Send prescriptions to pharmacy system
Health insurance verification API	Interface	Check patient's insurance status

Function Points Metric (Example: Hospital Management System)

Component	Count	Weight	Contribution to UFP
Inputs	4	4	$4 \times 4 = 16$
Outputs	3	5	$3 \times 5 = 15$
Inquiries	2	4	$2 \times 4 = 8$
Master Files	2	10	$2 \times 10 = 20$
Interfaces	2	7	$2 \times 7 = 14$

Total UFP = 16 + 15 + 8 + 20 + 14 = 73

Function Points Metric (Example: Hospital Management System)

Technical Factor Description	Rating	Reason for Rating
Backup and recovery needed/ data communicatoin	4	Patient and hospital data must be reliably backed up and recoverable in case of failure.
Distributed system	0	In a small hospital or clinic setup, the system might be standalone and not distributed.
Performance is critical	5	Real-time access to patient info and medical data is critical in emergencies.
Complex internal processing	4	Handling appointments, billing, medical history, and diagnostics involves complex logic.
Reusability	2	Some components like login and reporting can be reused, but many features are custom.
Installation ease	0	System is installed once on a local server by IT staff; no major complexity involved.
User-friendliness	3	Important for medical staff, though not necessarily highly intuitive in all parts.
Portability	0	The system is not expected to run across different OS or devices—used only on-site.
Maintainability	2	Occasional updates are needed, but it's not highly complex to maintain.
Concurrent users supported	4	Multiple staff members (nurses, doctors) use the system at the same time.
Security features	5	Strong security needed to protect sensitive patient data.
Third-party integration (pharmacy/insurance)	3	Needs to communicate with external services for prescriptions and insurance.
End-user training required	2	Staff need some basic training to use the interface and workflows.
Special interface requirements	0	No custom or device-specific interfaces are used in this basic version.

Sum of Technical Factor Ratings = **34**

TCF Formula: $TCF = 0.65 + (0.01 \times \text{Sum of Technical Factors})$

TCF = 0.65 + (0.01 × 34) = 0.65+0.34 = 0.99

Function Points Metric (Example: Hospital Management System)

$$\mathbf{FP = UFP \times TCF}$$

$$FP = 73 \times 0.99 = \mathbf{72.27}$$

This adjusted size can now be used for:

- **Estimating person-hours:**
  1 function point (FP)= 6 person-hours (industry benchmark),
  Total effort = 72.27 × 6 = 433.62 person-hours (1 person working for 433.62 hours)
- **Planning cost:**
  If the hourly cost is $50, then
  Total cost = 433.62 × 50 = $21,681
- **Comparing with other projects:**
  The project can be benchmarked against similar systems based on their FP size to evaluate efficiency, cost-effectiveness, and complexity.
- **Planning schedule:**
  If 3 developers work full-time (8 hours/day), then
  Total duration ≈ 433.62 / (3 × 8) = approx. 18.07 workdays

2. Algorithmic Models

- These are mathematical models used to compute cost, time, and resources.
- Unlike human judgment, these models remove bias and base decisions on structured formulas.
- However, they're only as good as the assumptions they rely on.

- Examples
  - Emperical Model- Based on observed project data and mathematical relationships to estimate effort, cost, and schedule.
  - Price S Model - Uses statistical estimation techniques.
  - Constructive Cost Model (COCOMO) - A well-known cost estimation model in software engineering.

Empirical Models

- A software estimation approach based on historical project data and observed outcomes.
- Uses mathematical formulas derived from analyzing many past projects.
- Helps estimate effort, schedule, and cost based on real-world metrics.
- **Key Characteristics:**
  - Data-driven: relies on actual project data
  - Often uses statistical or regression techniques
  - Can include probability distributions for uncertainty
  - Does not require complete technical specs early on

Intermediate COCOMO (Constructive Cost Model)

COCOMO is a model used to estimate the effort, time, and cost of a software project based on its size.

Used to estimate:

- Effort (person-months)
- Development time
- Project cost
- Effort estimation formula:

$$\textbf{Effort} = a \times (KDSI)^b \times EAF$$

- Where:
  - **KDSI(Kilo Delivered Source Instructions)** = Thousands of Lines of of finished, working code
  - **a, b** = Constants based on project type
  - **EAF** = Effort Adjustment Factor (from cost drivers)

Example: If you're developing a simple mobile game, the cost will be much lower than developing NASA's flight control software, because one is an organic project, while the other is embedded.

Intermediate COCOMO (Constructive Cost Model)

Identify the type of development mode:

- **Organic (2 – 50 KDSI)** – Small teams, straightforward projects.
- **Semi-detached (50 – 300 KDSI)** – Medium complexity, some teamwork.
- **Embedded (> 300 KDSI)** – High complexity, strict constraints (e.g., aerospace software).

4. Calculate nominal effort
   - Example: Straight forward product ("**organic mode**"):

     Nominal effort = 3.2 X $(KDSI)^{1.05}$ person-months

     Nominal effort = 3.2 X $(12)^{1.05}$ = 43 person-months

5. Multiply nominal value by 15 software development cost multipliers

Example: If you're developing a simple mobile game, the cost will be much lower than developing NASA's flight control software, because one is an organic project, while the other is embedded.

Intermediate COCOMO (Effort Adjustment Factor)

Product Complexity multipliers

Cost Drivers	Rating					
	Very Low	Low	Nominal	High	Very High	Extra High
Product attributes						
Required software reliablity	0.75	0.88	1.00	1.15	1.40	
Database size		0.94	1.00	1.08	1.16	
Product complexity	0.70	0.85	1.00	1.15	1.30	1.65
Computer attributes						
Execution time constraint			1.00	1.11	1.30	1.66
Main storage constraint			1.00	1.06	1.21	1.56
Virtual machine volatility*		0.87	1.00	1.15	1.30	
Computer turnaround time		0.87	1.00	1.07	1.15	
Personnel attributes						
Analyst capabilities	1.46	1.19	1.00	0.86	0.71	
Applications experience	1.29	1.13	1.00	0.91	0.82	
Programmer capability	1.42	1.17	1.00	0.86	0.70	
Virtual machine experience*	1.21	1.10	1.00	0.90		
Programming language experience	1.14	1.07	1.00	0.95		
Project attributes						
Use of modern programming practices	1.24	1.10	1.00	0.91	0.82	
Use of software tools	1.24	1.10	1.00	0.91	0.83	
Required development schedule	1.23	1.08	1.00	1.04	1.10	

*For a given software product, the underlying virtual machine is the complex of hardware and software (operating system, database management system) it calls on to accomplish its task.

Example – Hospital Management System

- **Project Type:** Organic
- **Estimated Size:** 25 KDIS
- **EAF (from cost drivers):**

Cost Driver	Rating	Multiplier
Required Software Reliability	Nominal	1.00
Product Complexity	Nominal	1.00
Programmer Capability	High	0.86
Use of Software Tools	Nominal	1.00
Execution Time Constraint	Very High	1.30

To

- **Constants for Organic:** a = 3.2, b = 1.05

  Nominal effort = 3.2 X $(25)^{1.05}$ X1.12

  ≈ 3.2 × 27.15 × 1.12 = 97.5 person-months (One person working full-time for 97.5 months)

Person-hours: 97.5 PM × 160 hrs (typical full-time work in 1 month) = 15,600 hrs
Cost: 15,600 × $50/hr = $780,000
Schedule (with 5 developers):
15,600 / (5 × 160) ≈ 19.5 months

Intermediate COCOMO

Example: "Embedded" communications processing software for electronic funds transfer network; high reliability, performance, development schedule, and interface requirements."

1. Complex ("embedded") mode
2. Estimated to be 10,000 KDSI. (i.e., 10 million lines of code)
3. Nominal effort = 2.8 X $(10000)^{1.20}$ = 63,096 person-months
4. Product of complexity multipliers = 1.30

Effort estimate: 1.30 X 63,096 = 229,669.44 person-months

Intermediate COCOMO

Effort Multipliers.

Cost Drivers	Situation	Rating	Effort Multiplier
Required software reliability	Serious financial consequences of software fault	High	1.15
Database size	20,000 bytes	Low	0.94
Product complexity	Communications processing	Very high	1.30
Execution time constraint	Will use 70% of available time	High	1.11
Main storage constraint	45K of 64K store (70%)	High	1.06
Virtrual machine volatility	Based on commercial microprocessor hardware	Nominal	1.00
Computer turnaround time	Two hour average turnaround time	Nominal	1.00
Analyst capabilities	Good senior analysts	High	0.86
Applications experience	Three years	Nominal	1.00
Programmer capability	Good senior programmers	High	0.86
Virtual machines experience	Six months	Low	1.10
Programming language experience	Twelve months	Nominal	1.00
Use of modern programming practices	Most techniques in use over one year	High	0.91
Use of software tools	At basic minicomputer tool level	Low	1.10
Required development schedule	Nine months	Nominal	1.00

Intermediate COCOMO

- Estimated effort for project (e.g., 229,669.44 person-months) is then used as input for additional project estimates:
  - Dollar costs.
  - Development schedules.
  - Phase and activity distributions.
  - Computer costs.
  - Annual maintenance costs.
  - Other related items.

- Accuracy of COCOMO estimates
  - According to Boehm: predicted values lie within 20% of the actual value, about 68% of time
  - Is that good enough?

4. Expert judgement

- An expert in software development as well as in the application domain makes an estimate based on previous experience of similar projects.

Expert judgement – pros & cons

- Advantages:
  - Relatively cheap estimation method.
  - Takes relatively little time and effort
  - Can be applied early in the development cycle
  - Can be successful if experts have direct experience of similar systems
- Disadvantages:
  - Rather subjective
  - Depends on experience and judgment
  - Cannot be used if no suitable experts available
  - Assumes experts have dealt with similar systems
  - Assumes they all have reliable data available

5. Sum of the parts

- Makes use of work breakdown structure
- Total effort estimate is the sum of estimates for individual tasks
- Appropriate level of detail (granularity) is important
  - too much detail takes too much time and introduces more error
  - insufficient detail means more difficult to assign tasks
- Must make allowance for overheads and tasks such as testing and documentation

Determining
1. Work breakdown
2. Durations
3. Dependencies

The hard part of project estimation

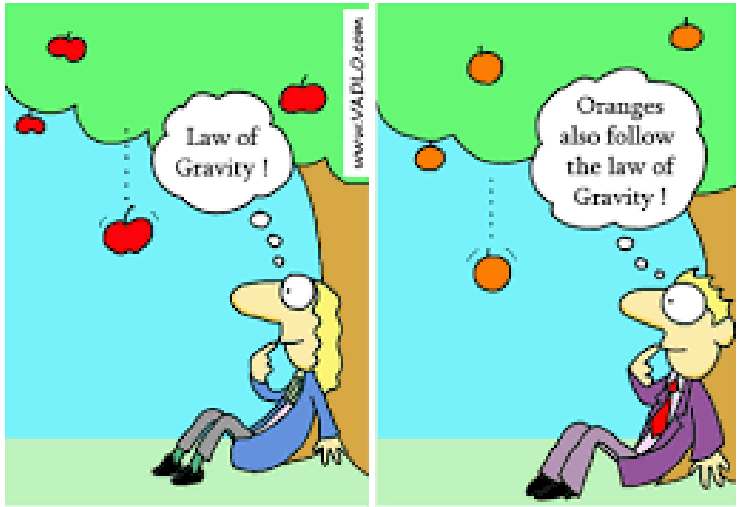
Task ID	Task Description	Duration (in weeks)	Predecessors
A	Initiate project	0	--
B	R & D product design	6	A
C	Plan market research	2	A
D	Routing (engineering)	3	B
E	Build prototype	5	B
F	Prepare brochure	3	B
G	Prepare cost estimates	2	D
H	Product testing	3	E
I	Market survey	4	C, F
J	Pricing and demand forecast	2	I
K	Final report	2	G, H, J

Sum of the parts example

		Week							
Task	Hours	5	6	7	8	9	10	11	12
Research	1	----	----						
Draw plan	2		----						
Obtain tools	1		----						
Set-up work bench	0.5		----						
Buy balsa	1			--					
Testing equipment	0.5			--					
Construct 'beams'	2				----				
Construct roadway	2					----			
Build structure	3						----		
Check clearances	0.5						--		
Complete structure	3							----	
Final test	1								----
Total	17.5								

6. Estimation by analogy*

- Compare current project to similar project(s) already undertaken
- Estimate how many times bigger or smaller the current project is compared with others



High Impact Paper | Low Impact Paper

*See for example, Martin Shepperd, Chris Schofield and Barbara Kitchenham, *Effort Estimation Using Analogy*, Proceedings of ICSE-18, IEEE, 1996

Using analogy – pros & cons

Advantages:

- Systematic, fairly fast
- OK if sufficient historical data available
- Can be applied early in the development cycle

Disadvantages:

- Have to determine set of characteristics suitable for classifying systems
- Requires a database containing systematically maintained historical size cost data.
- Cannot be used if no comparable projects have ever been tackled, or if no suitable historical data is available

Choosing estimation methods

Which is easiest to apply?

Which can be applied earliest in the system development life cycle (SDLC)?

What assumptions does each make?

Do I have enough historic data?

Will I need to re-calibrate for tool, developer experience, environment, etc.

How many times will I use this method?
Can I improve my estimates over time?

Six approaches to project estimation

1. Function point analysis
2. Algorithmic cost models
3. Component matrix
4. Expert judgement
5. Sum of the parts
6. Estimation by analogy

But what aspects of a project are these taking into account?

- Development, Delivery, Transformation, Operations, Maintenance?
- All can be inaccurately measured

END