INFO5990: Professional Practice in IT

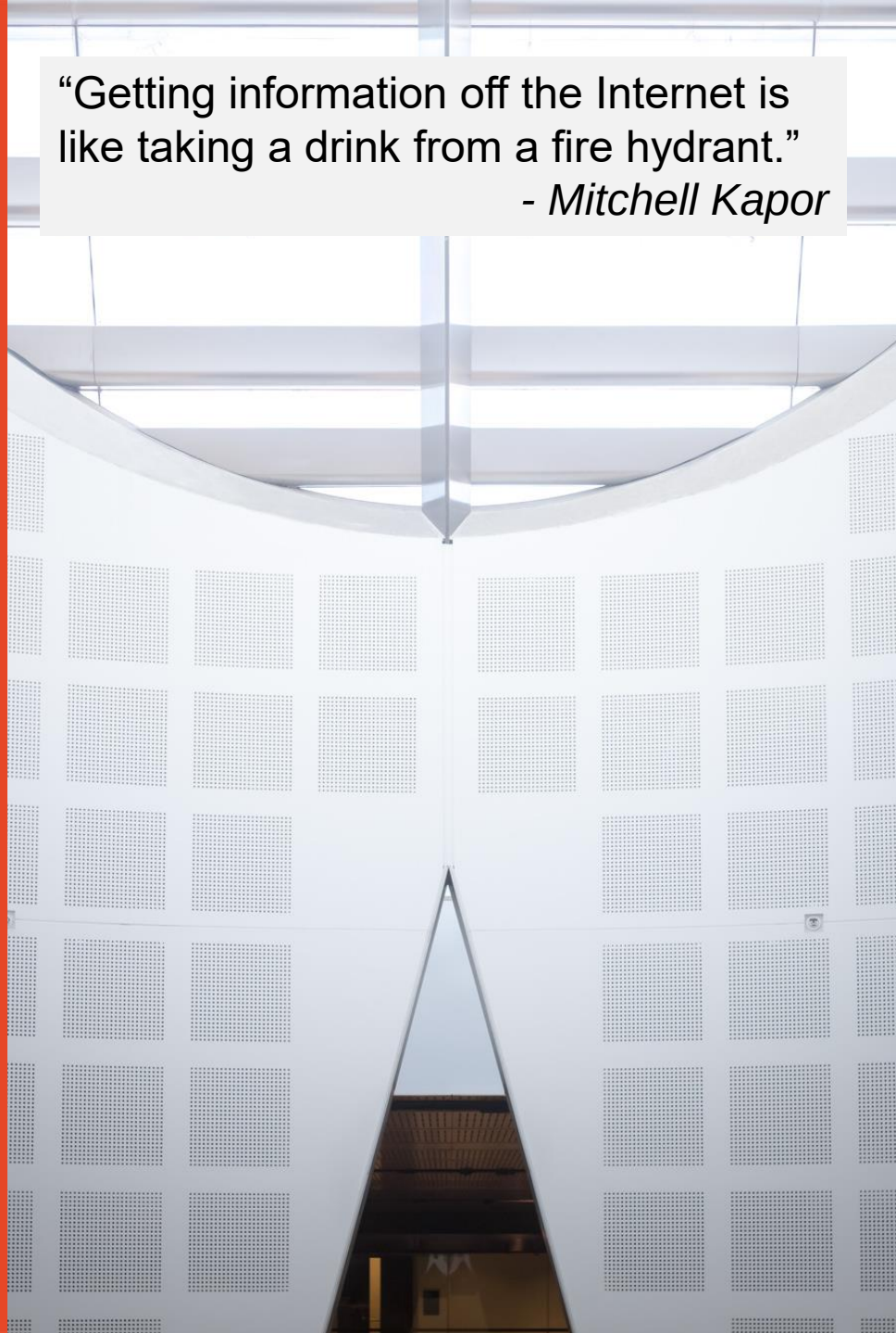
Week 5:

- Information
- Research
- Estimation

Dr. Reza Hoseiny

School of Computer Science

THE UNIVERSITY OF SYDNEY



Quick Overview of Today

Quiz Review

Part A: Information

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

Part B: Research

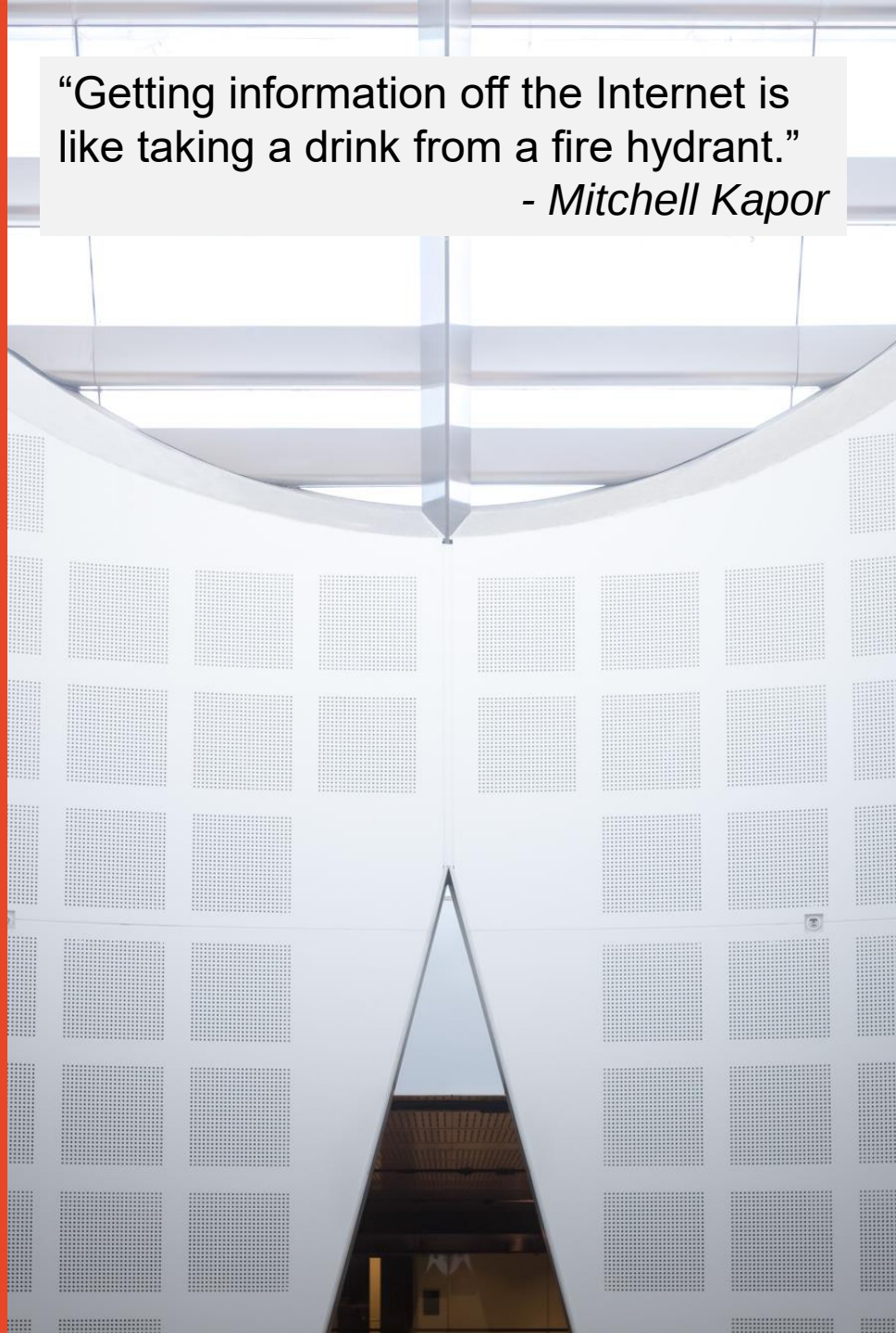
- Statistics
- Types of research
- Business analytics

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?
  - Where his wife was born?
  - As the new Vice-Chancellor, what does "success" of his role looks like?
  - What his mother-in-law gets real joy from reading?
  - When did he move to Australia?
  - What his opinion is on the current the value of the Paralympics?

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.

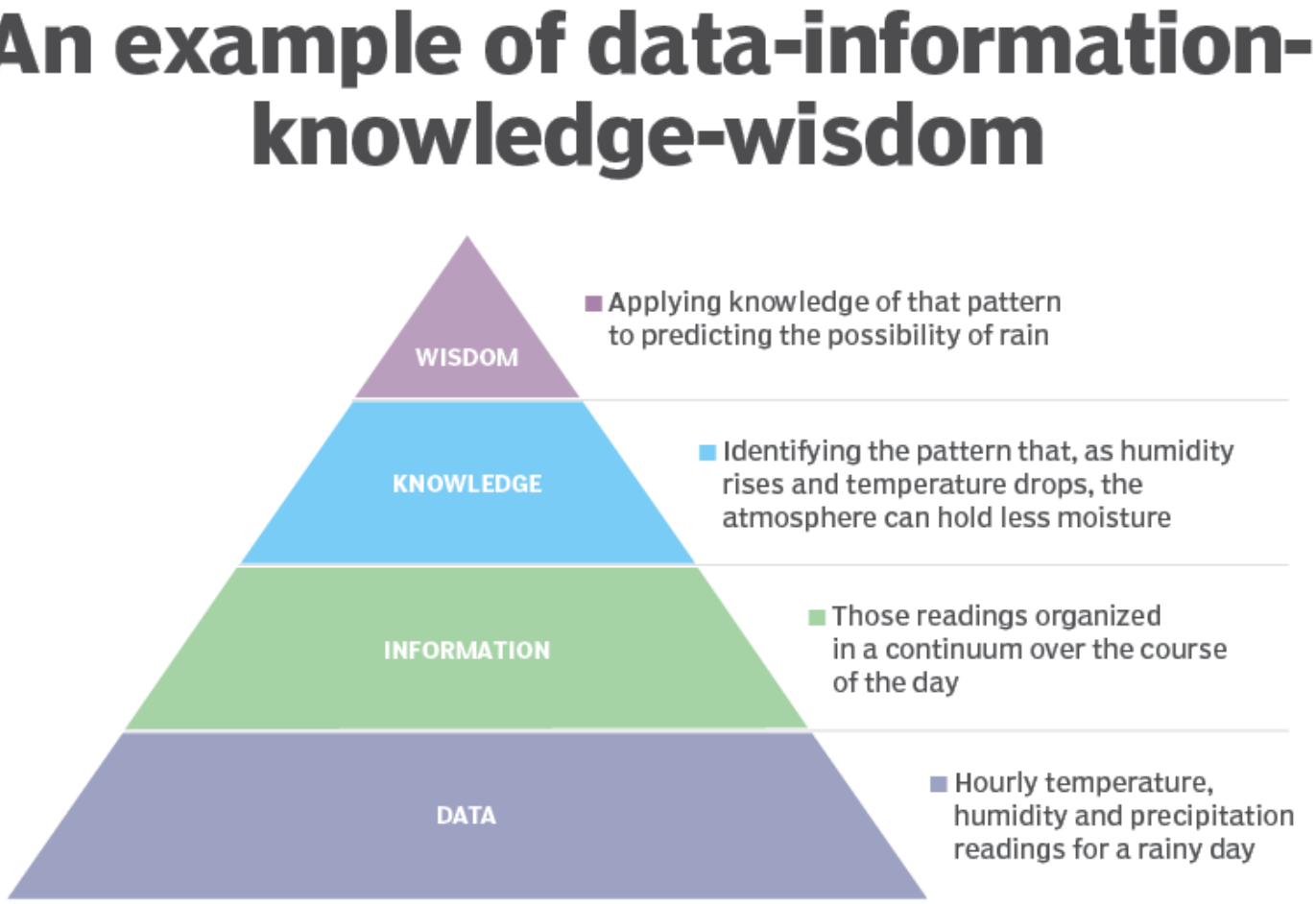
The Best Small Business CRM Software for 2021

The Best Marketing Automation Software for 2021

The Best Email Marketing Software for 2021

TOP PICKS

Data – Information – Knowledge - Wisdom



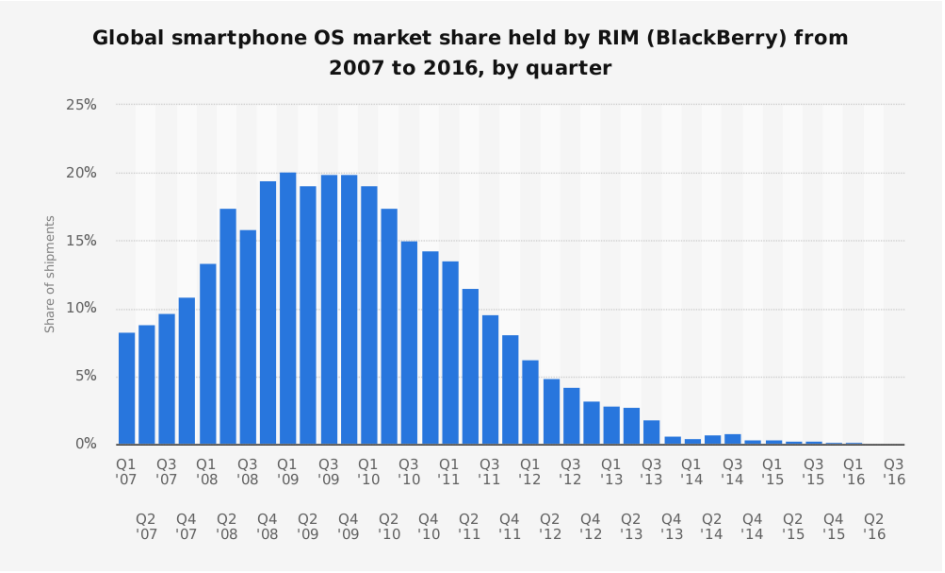
THE UNIVERSITY OF SYDNEY

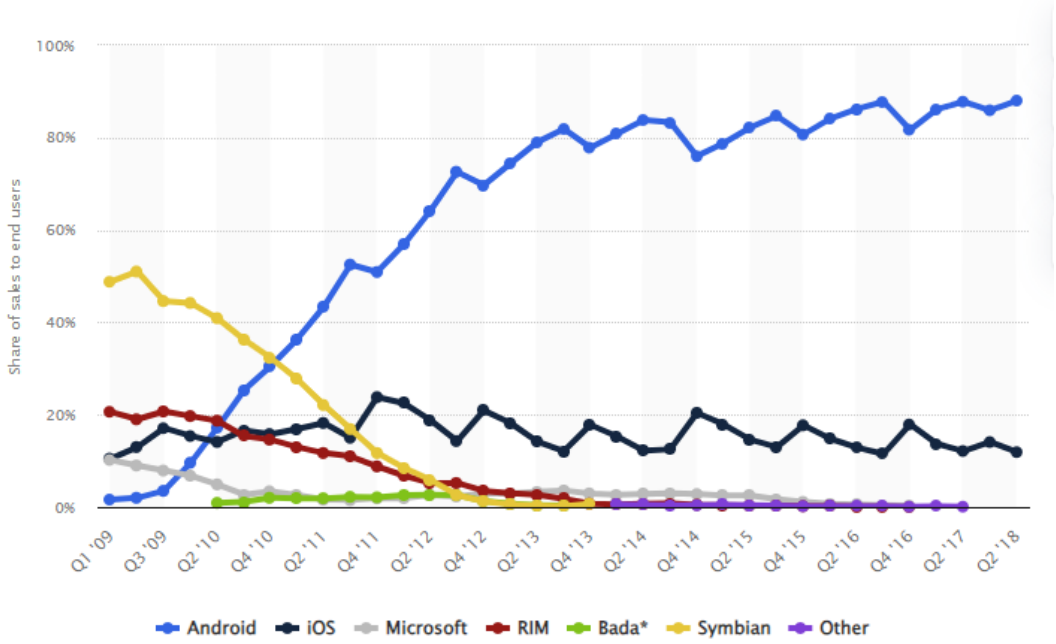
Source: https://www.techtarget.com/searchdatamanagement/definition/information

What level of information / knowledge is needed?

Immediate: narrow technical	• Ability to solve a problem *now* • e.g. *How do I do a merge into a git reposity?*
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. *What are the capabilities of GSuite vs O365?*
Possibilities: business value	• Does something add value? • e.g. *Might implementing a CRM add greater value than the cost?*
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, but not valid
  - 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

- Are the following reliable? valid?
  - https://www.wikipedia.org/
  - https://www.w3schools.com/
  - https://stackoverflow.com/
  - https://www.gartner.com/
  - https://dl.acm.org/

- **Is the following valid?**

  "My number one issue with GSuite is the lack of native Outlook connectivity for full email, calendar and contact syncing. There is the GSuite sync tool, but I've found it to be buggy and not 100% reliable. This is especially important when working across multiple devices, and needing consistent contact and calendar data on all devices."

  – https://www.wpbeginner.com/opinion/g-suite-vs-office-365-comparison-which-one-is-better/

- How do you know?

- A fun site to explore: https://www.snopes.com/

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 do a merge into a git repository?*
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. *What are the capabilities of GSuite vs O365?*
Possibilities: business value	• Does something add value? • e.g. *Might implementing a CRM add greater value than the cost?*
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?*

Team Report - finding sources of information

- The internet is a source of many different kinds of information. However, for your Team Report you need to find information that is sufficiently trustworthy for the Internal Investment Board (IIB) to use for its business case. Here are some tips to help you start looking for that material.

- For your Team Report you must select a company and an enterprise system. A good place to start selecting a company for your report is https://libguides.library.usyd.edu.au/market_intelligence. This links to an external site., specifically, the 'ASX-Listed Companies' and using the 'Company 360' database to find information specific to a company and 'Industry information' - 'IBISWorld' to find information about different industries.

- Databases such as this one, that are accessible through our Library, are good places to start conducting research because they provide credible resources such as journals, books, reference titles, where information is likely to come from experts, and also newspapers, magazines and much more. A library database will also clearly indicates the author and source. Library databases also offer options like related terms, search options to broaden or narrow the search, and other organisational tools that support your research.

- You can also use Google Scholar to find some, but not all, relevant information. Google Scholar cannot access all the information that is in the Library databases (because some of them are behind a paywall and are funded by the University), so do not rely on Google Scholar exclusively.

See - https://canvas.sydney.edu.au/courses/59195/modules for more information

Evaluating sources of information

'REVIEW' CRITERIA		'REVIEW' QUESTIONS
R	Relevance	Is the information in the source directly relevant to the requirements for your report?
E	Expertise of Author	1. What are their qualifications? 2. Are they writing in their area of expertise? 3. Are they cited by other authors in the field?
V	Viewpoint of Author/Organisation	1. Does the author have any personal or professional affiliations that may bias their work (e.g. a blog written by an IT supplier)? 2. Has the information been sponsored by an organisation who might gain from publishing the information (as above)? 3. What is the purpose of the source – to inform, persuade, or entertain?
I	Intended Audience	Is the source aimed at the general public, professionals, or scholarly audience?
E	Evidence	1. Are statements supported by evidence? - 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 …*, I read a blog that said) 2. Are the references to the evidence correct? 3. Has the source been subjected to peer review?
W	When Published	1. When was the source published? 2. Have significant developments been made in the subject area since the source was published?

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

Team Report - 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.

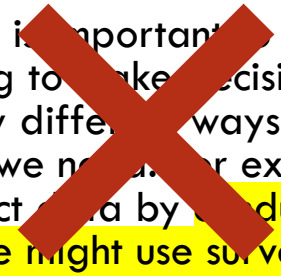
Plagiarism

- 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 a 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.

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 m[illegible] decision[illegible]ere are many differe[illegible]ays [illegible]btain the data that we need. F[illegible]mple, we might collect data usin[illegible]fferent approaches. Different peo[illegible] su[illegible]s scientists, historians, [illegible]end to [illegible] different approaches. These might include experiments, surveys, etc.

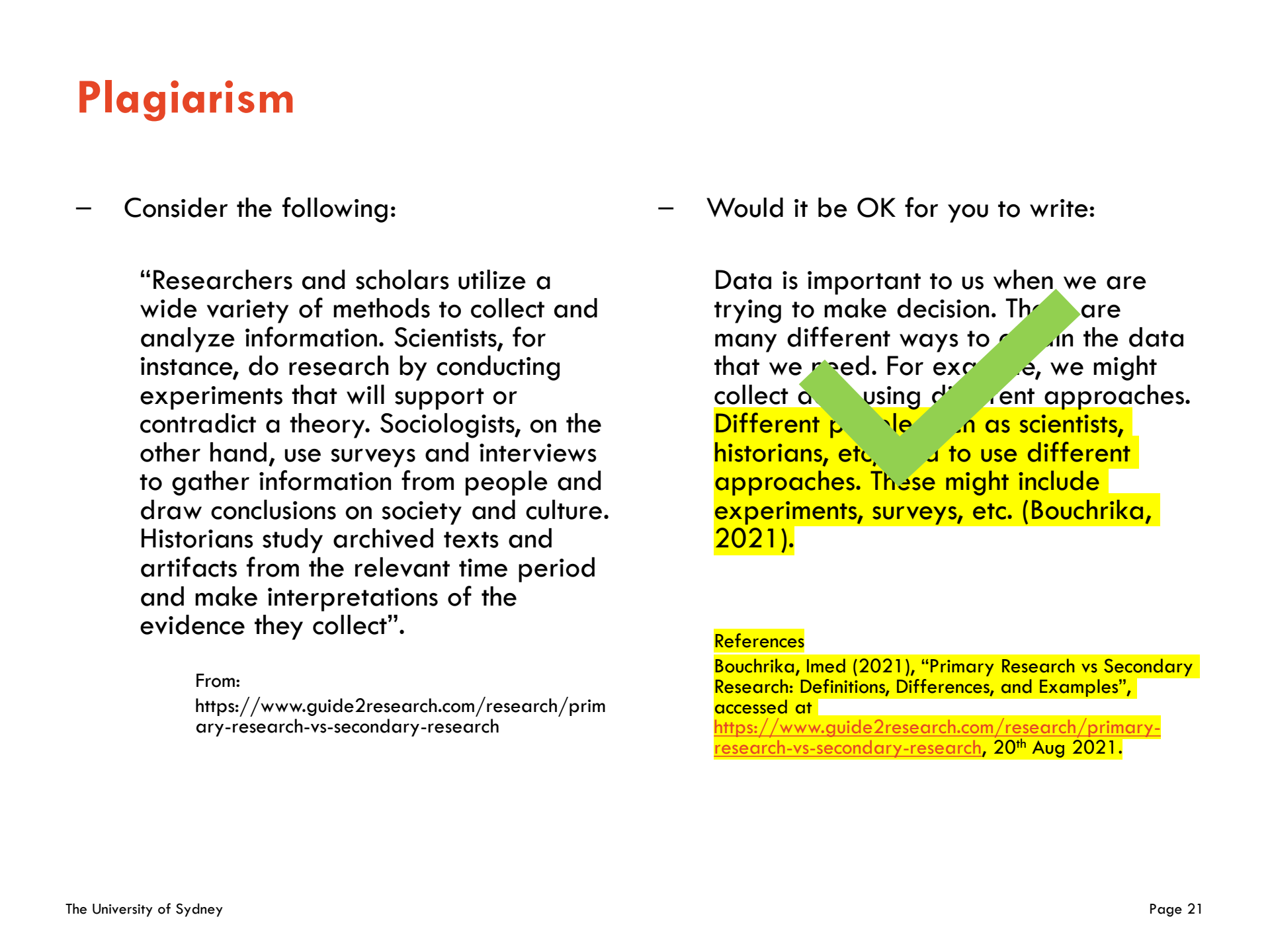
Plagiarism

- 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. Th[illegible] are many different ways to [illegible]in the data that we [illegible]eed. For exa[illegible]e, we might collect a[illegible] using di[illegible]ent approaches. Different p[illegible]le[illegible]n as scientists, historians, etc[illegible] 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.

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, [we could] study archived texts and artifacts" (Bouchrika, 2021).

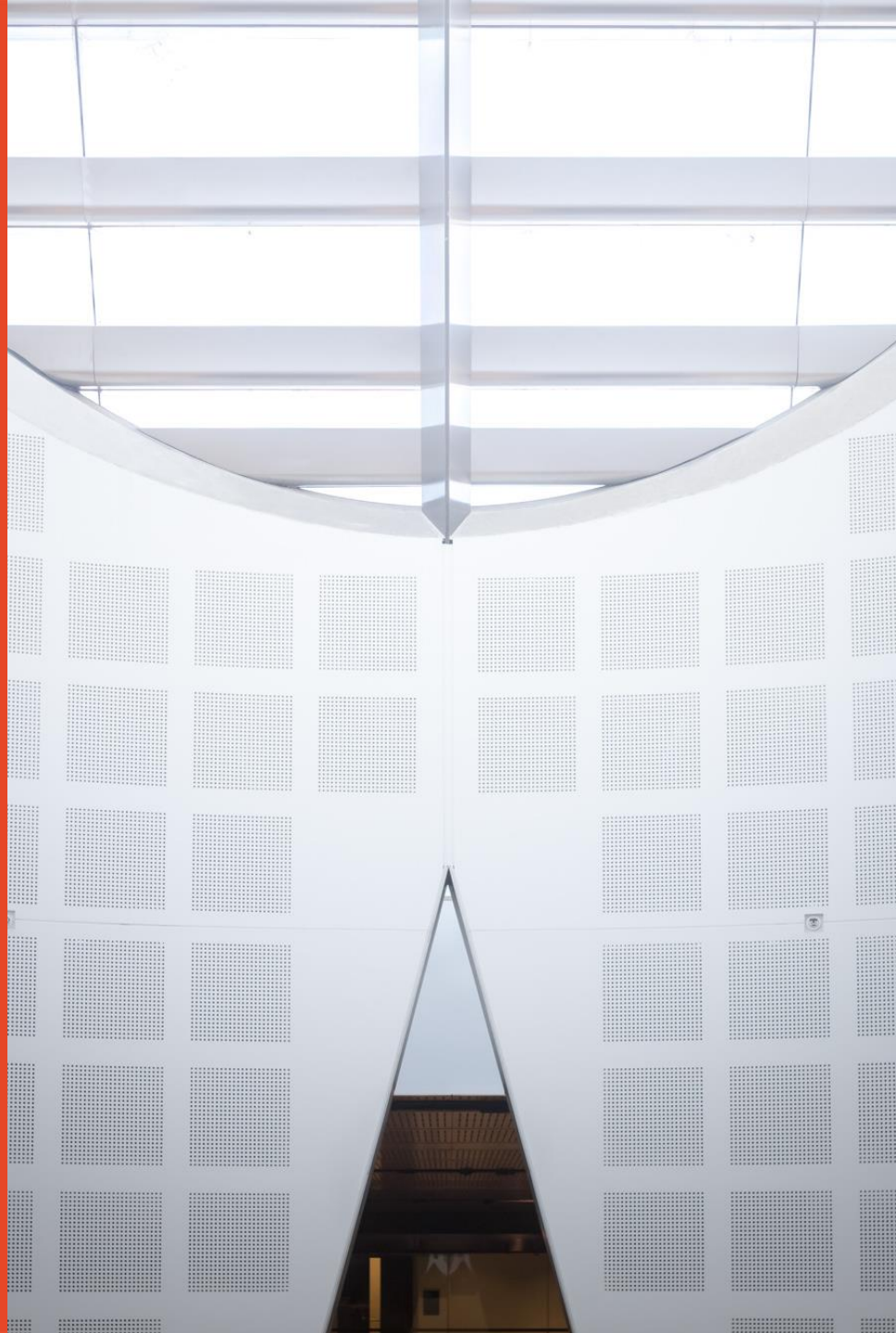
References

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.

INFO5990: Professional Practice in IT

Week 5

Part B: Research



THE UNIVERSITY OF SYDNEY

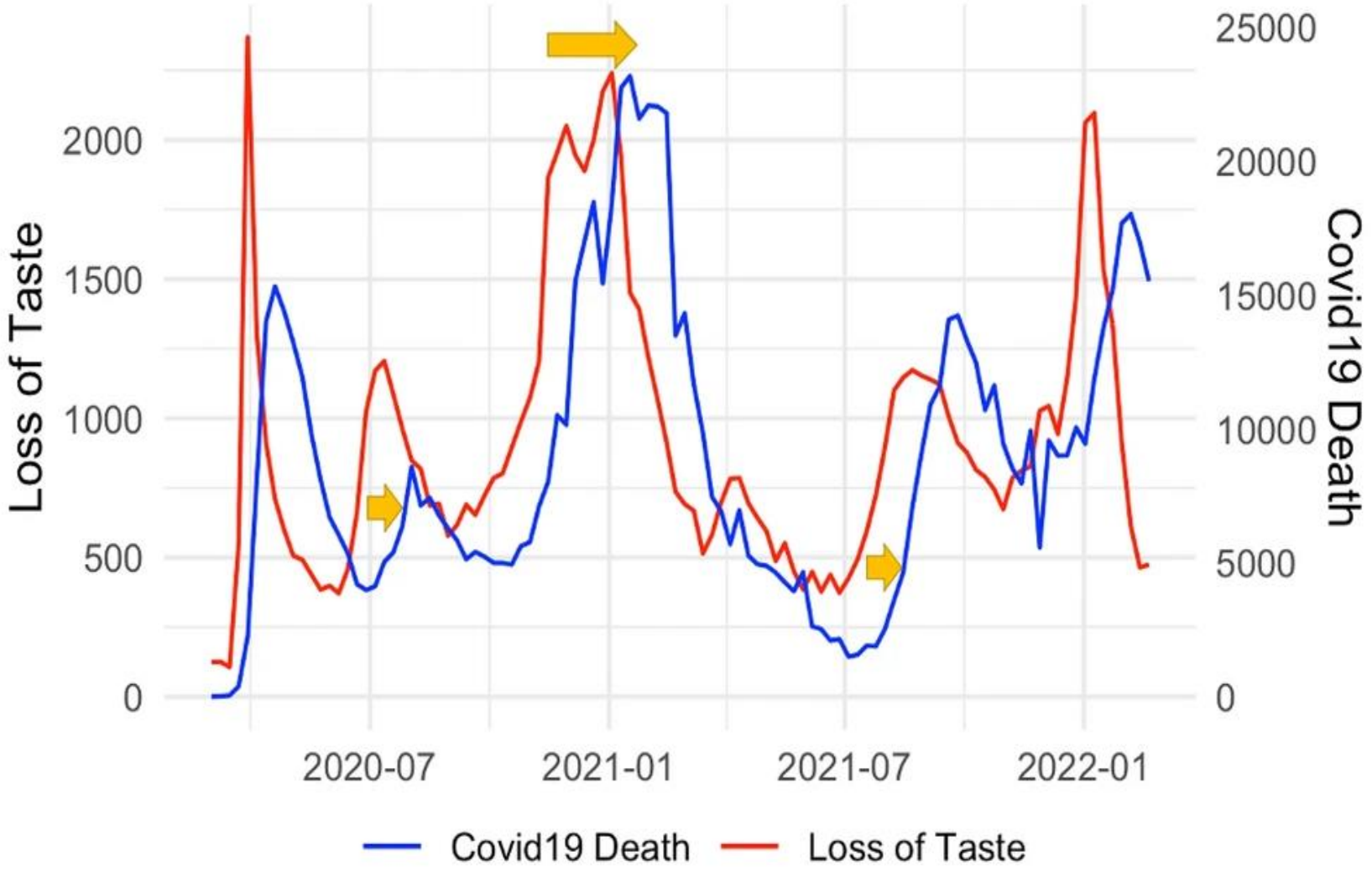
Research - why

- **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.

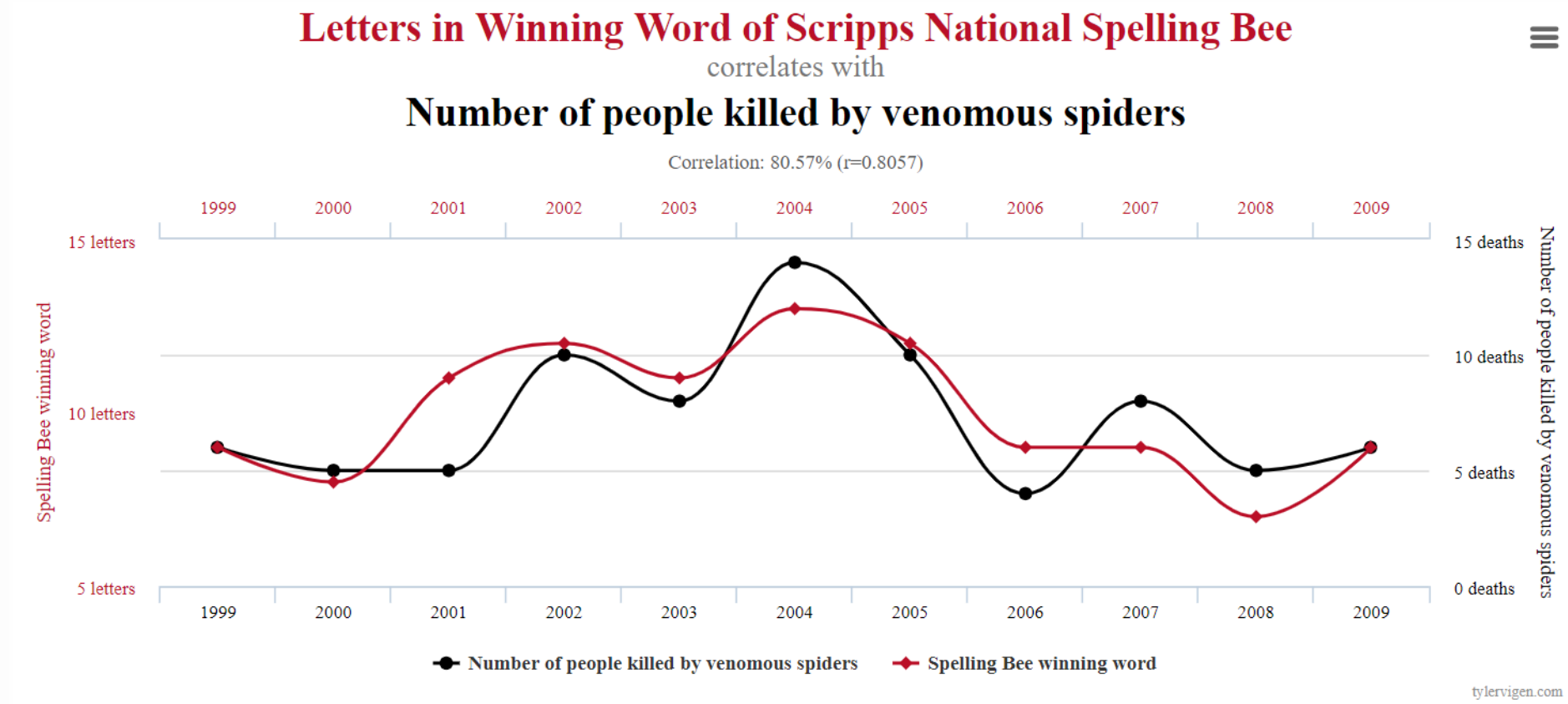
Knowing about statistics

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



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

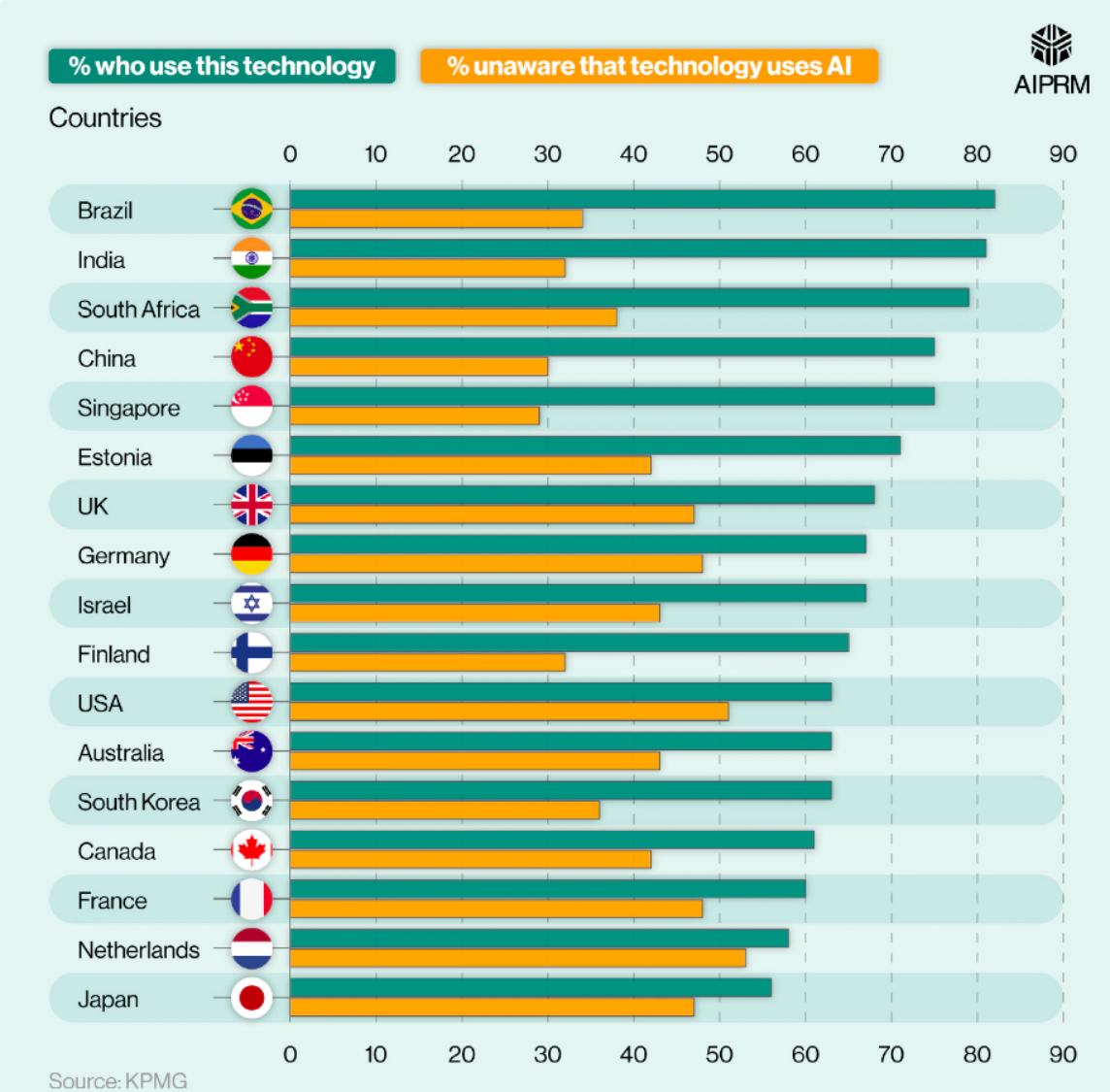
Knowing about statistics



Data sources: National Spelling Bee and Centers for Disease Control & Prevention

Knowing about statistics

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



Knowing about statistics

Common mistakes

- Data ≠ Insight
- Correlation ≠ Causation
- Poorly framed questions
- Sample sizes and statistical significance
- Confounding factors

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/

Example: Remote Laboratory adoption

Industry = Engineering Education

Context = Laboratory experiments using real equipment

- E.g. Civil Engineering students do experiments where they put a load onto a beam and see how much it deforms.
- But access to the lab is only during restricted times, and so the time available to do the experiment is limited.
- And many institutions don't have the necessary equipment.

Technology = Internet enabling of access

- Connect sensors, actuators and cameras to the equipment, so it can be accessed at any time
- Include a booking / queuing system and data logging.

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. MTBF; Bug reports/tickets

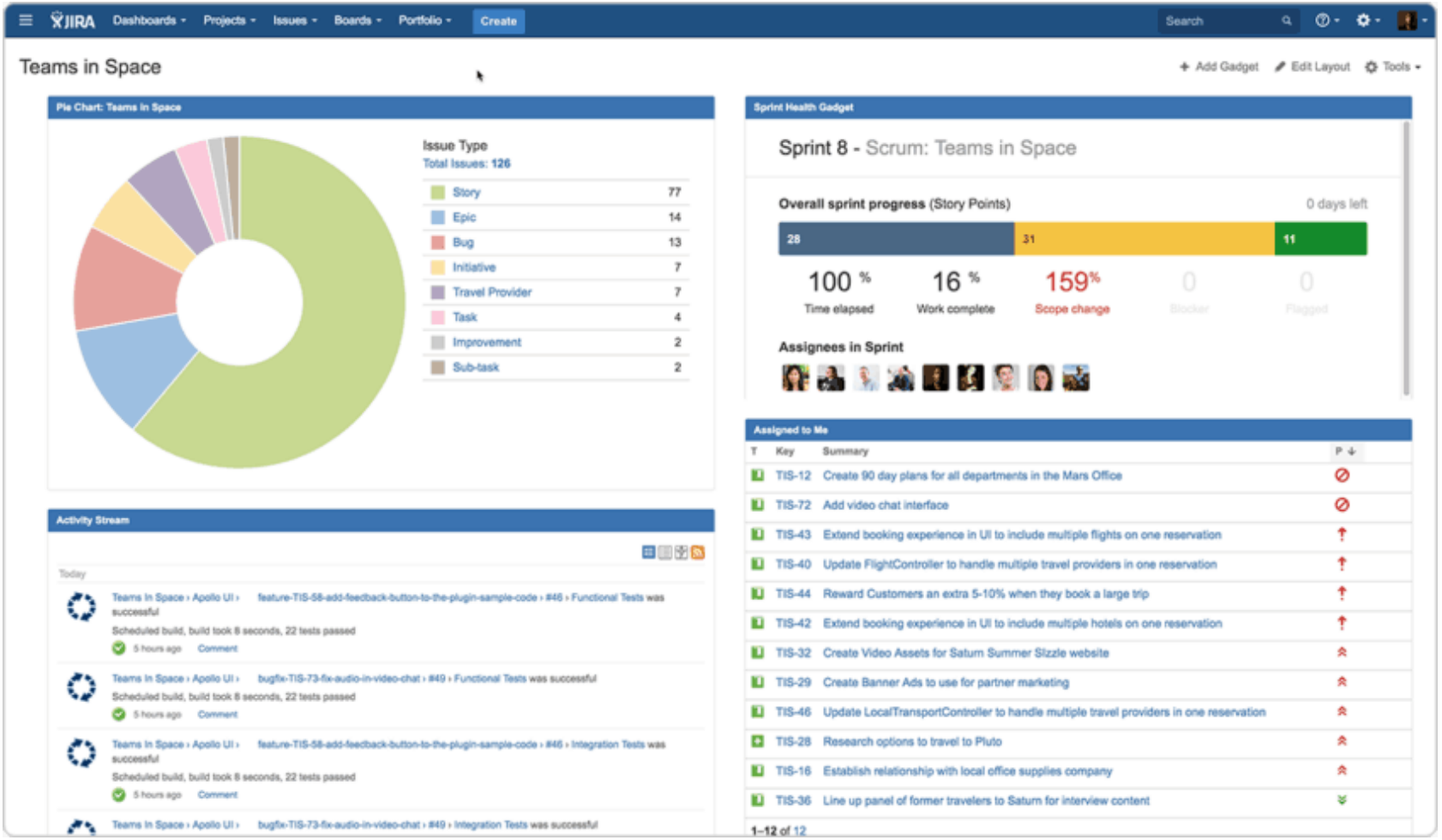
Business Operations

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

But remember: data ≠ insight

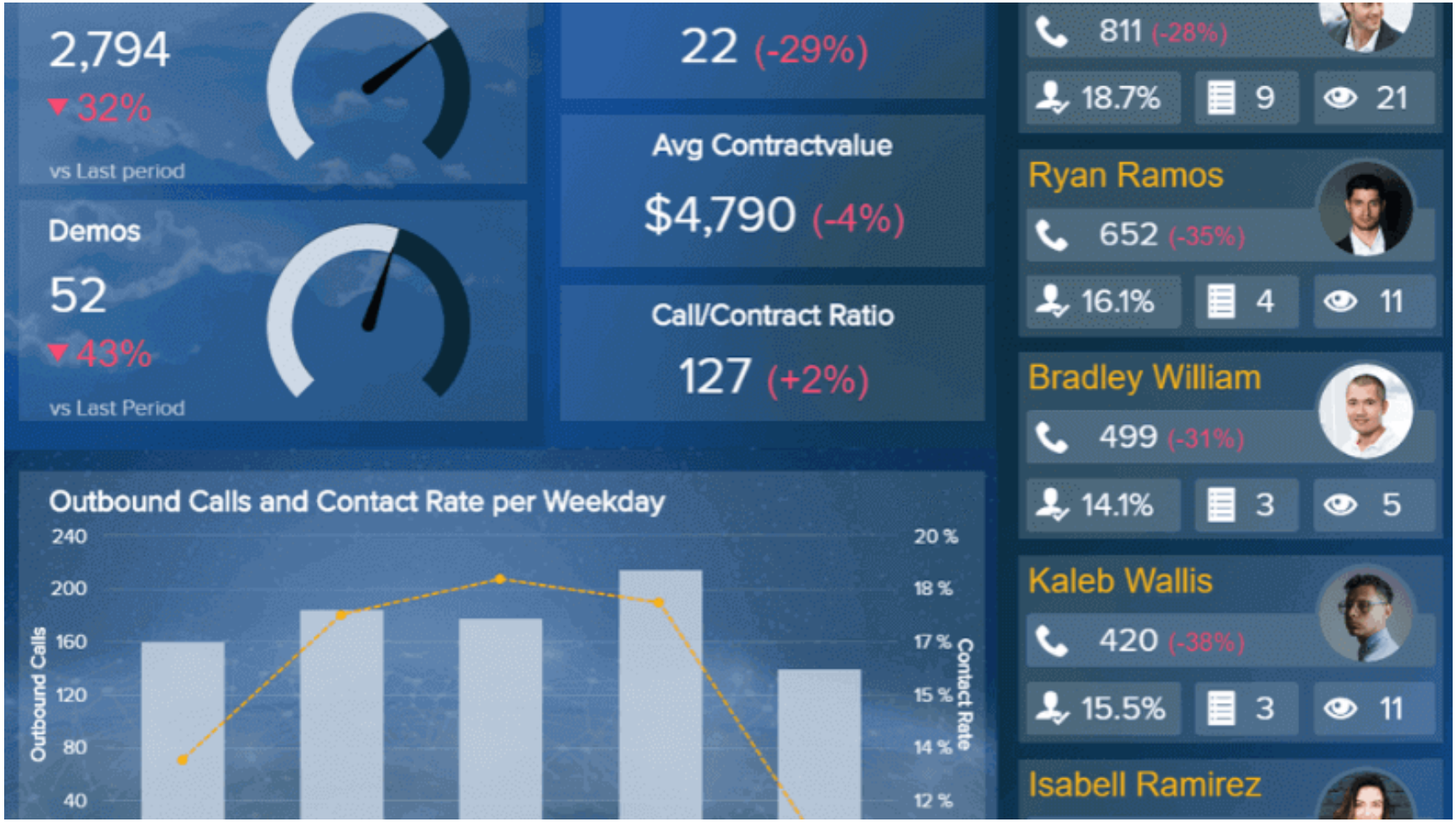
Business Analytics: Tools

Source: https://idalko.com/jira-dashboards/



Business Analytics: Example CRM Dashboard

https://www.datapine.com/blog/best-crm-dashboard-report-examples/
Also:
https://fitsmallbusiness.com/what-is-a-crm-dashboard/



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

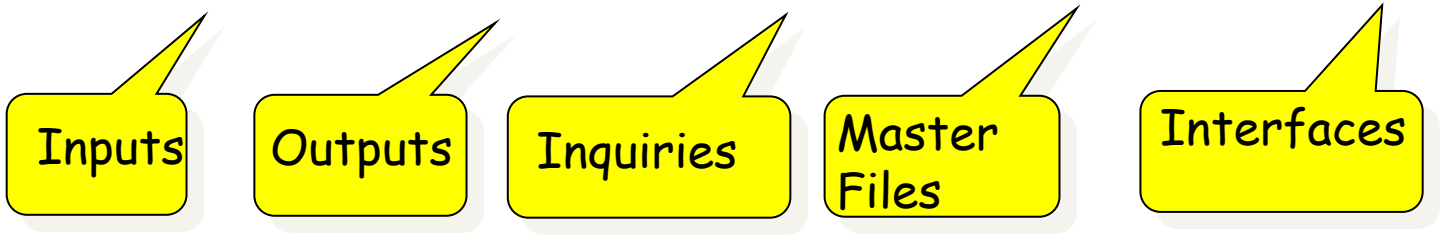
Six approaches to project estimation

- Function point analysis (not common)
- Algorithmic cost models (not common)
- Component matrix (not common)
- Expert judgement
- Sum of the parts
- Estimation by analogy

1. Function Points Metric

- Better medium-scale computer systems estimation metric.
- For any product, size in "function points":

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



- Oversimplification of a 3-step process…..

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

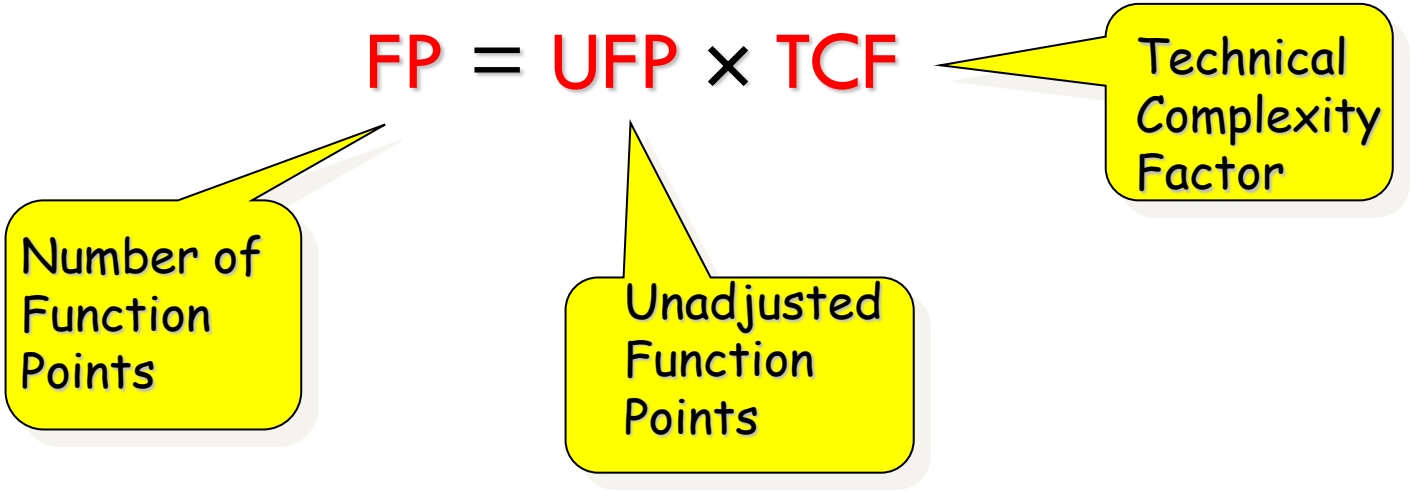
- 2. **Compute technical complexity factor (TCF)**
  - Assign value from 0 ("not present") to 5 ("strong influence throughout") to each of 14 factors.
  - Add 14 numbers = total degree of influence (DI)

    $$TCF = 0.65 + 0.01 \times DI$$
  - 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/

2. Algorithmic Models

- Used as input to compute cost, duration
  - Algorithmic model unbiased and superior to expert opinion.
  - However, estimates only as good as underlying assumptions
- Examples
  - SLIM Model
  - Price S Model
  - Constructive Cost Model (COCOMO)

Empirical Models

- A mathematical function that mimics some trend seen in observed data
- Has no underlying theory to help explain the behaviour
- Determined mathematically from historical data using "regression"
- Can be used to ***predict***, but not ***explain*** the behaviour of a system.

Intermediate COCOMO

- 1. Estimate length of product in KDSI
- 2. Estimate product development mode
  - (organic, semidetached, embedded)
- 3. 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

- 4. Multiply nominal value by 15 software development cost multipliers

Intermediate COCOMO

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.

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.
3. Nominal effort = 2.8 X $(10)^{1.20}$ = 44 person-months
4. Product of complexity multipliers = 1.30

   Effort estimate: 1.30 X 44 = 57.2 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., 57.2 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 (see work breakdown structure, Lec 6A)
- Must make allowance for overheads and tasks such as testing and documentation

The hard part of project estimation

Determining
1. Work breakdown
2. Durations
3. Dependencies

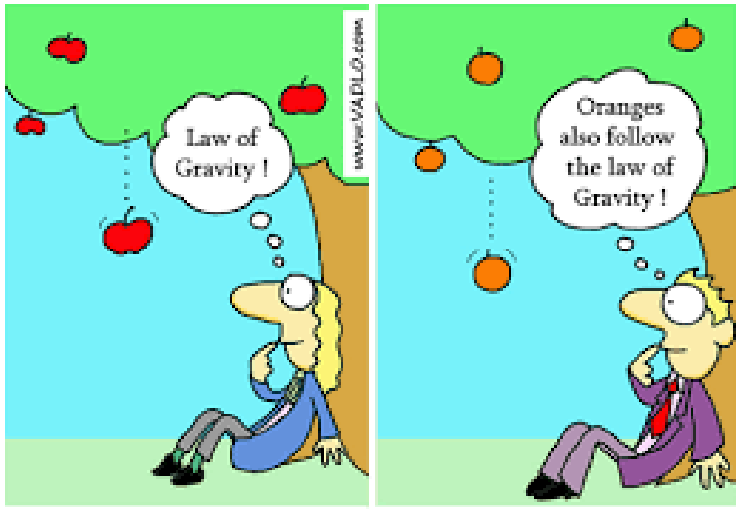
Task ID	Task Description	Duration (in weeks)	Predecessors
A	Initiate project	0	--
B	R & D product design	6	A
C	Plan market research	2	A
D	Routing (engineering)	3	B
E	Build prototype	5	B
F	Prepare brochure	3	B
G	Prepare cost estimates	2	D
H	Product testing	3	E
I	Market survey	4	C, F
J	Pricing and demand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

Simple Example

- Basic Information System for a Veterinarian
  - Records:
    - Customer information
      - Name, address, …
      - Pet(s): names, types, etc.
    - Visit details
      - Pet, treatment, payment details, …

Veterinarian Information System

- Function point calculation

Transactions	Input Data Elements	Entities Referenced	Output Data Elements
D1000: Enter/Edit job records	8	4	7
D4000: Enter hours from time sheets	6	5	4
D5000: Post timesheet data	0	4	5
M1000: Enter/edit client details	7	1	0
M2000: Enter/edit materials	7	4	7
R1000: Job progress report	0	3	5
R2000: Timesheet summary	0	5	7
Etc.	…	…	…
Totals	Ni	Ne	No

- Unadjusted Function Points (UFP)

  = WiNi + WeNe + WoNo

  = 0.58Ni + 1.66Ne + 0.26No

  = 112

Veterinarian Information System

- Final determination of effort
  - UFP = 112
  - Complexity = 0.65 + 0.005x44
    = 0.87
    - (ranges from 0.65 .. 1.15)
  - Adjusted FP = 0.87 x 112 = 97.4

No	Complexity Factor	Score 0-5
1	Data communications	4
2	Distributed system	3
3	Performance critical	0
4	Heavily used configuration	5
5	Transaction rates	2
6	Online data entry	4
7	Designed for end-user efficiency	5
8	Online update integrity	3
9	Complexity of processing	0
10	Usable in other applications	0
11	Ease of installation	2
12	Ease of operation	5
13	Multiple sites	0
14	Design to facilitate change	3
15	Requirements of other systems	0
16	Security privacy and auditability	4
17	User training needs	2
18	Direct use by third parties	0
19	Need for documentation	2
20	Any other special requirements	0
	Total	44

See:
https://www.researchgate.net/publication/274638632_What_Is_the_Cost_of_One_IFPUG_Method_Function_Point_-_Case_Study

Veterinarian Information System

- Converting from FPs to SLOC
  - Numerous factors
    - Language
    - Experience
    - Development environment
  - But there are "default" figures

Language	SLOC per FP
C++	53
COBOL	107
Java	46
SQL	13

  - 97.4 FP, using Java → 4.48 kSLOC

- And then to effort
  - Using Boehm's model to determine duration.
    - Assuming that this was an e-commerce system:
    - Effort = Productivity * KSLOC $^{\text{Penalty}}$

      $= 3.08 * 4.48^{1.030} = 14.4$ Person-Months

  - ??????

Discussion

- How reliable is this estimate?
- What factors might make it less/more reliable?
- What can we do to make the estimates more reliable?

- See:
  - https://www.isbsg.org/
- And read:
  - https://www.isbsg.org/wp-content/uploads/2020/01/Software-Cost-Estimate-for-a-Large-Trading-Company-1.pdf
  - 8,500 FP project
  - Likely time = 65,000 hours (~40 person years)
  - Likely cost = 5.2 million Euro

END