CHAPTER 1: INTRODUCTION TO DATA VISUALIZATION

STA191 – Fundamentals of Data Visualization Week 1 · Two sessions (1 h 50 min each) · Revised teaching note (Suriyati). Draft v2 — for review.

What this chapter does

This is the "why" chapter. Students don't build a single chart here — they build the *mindset* that makes every chart in later chapters worth making. Everything that follows (chart types in Chapter 2, Google Sheets and Power BI in Chapter 4, the case study in Chapter 5) rests on two ideas introduced here: **know your purpose** and **know your audience**. If students leave Week 1 understanding those two things, the rest of the course is technique.

Chapter learning outcomes

By the end of this chapter, students should be able to:

1. **Explain** what data visualization is and where it sits inside the data analysis process.
2. **Distinguish** exploratory from explanatory data analysis, and identify which one a given situation calls for.
3. **Ask** the three context questions (audience, message, mechanism) before making any visual.
4. **Describe** the steps taken when meeting a new dataset for the first time.
5. **Justify** why we visualize data — to communicate, to turn data into information, and to show evidence.

(These map to CLO1/CLO2 at the awareness level — the doing comes later.)

How Week 1 is split

	Focus	Main content
Session 1	Course briefing + *Why visualize, and for whom?*	Admin block → §1.1 What is data viz → §1.2 Exploratory vs explanatory → §1.3 Three context questions → activity
Session 2	*From dataset to evidence*	Recap → §1.4 Meeting a dataset (live walkthrough) → §1.5 Scholarly perspectives → §1.6 Three reasons to visualize → discussion + preview

SESSION 1 — Course Briefing + Why Visualize, and For Whom?

0. Course briefing (first class only)

Week 1's first session is the students' first contact with the course. Spend the opening block orienting them before any content — they need to know what STA191 is, how they'll be graded, and what's expected. Keep it brisk and welcoming; the concepts start straight after.

0.1 What this course is

STA191 – **Fundamentals of Data Visualization** (*Asas Visualisasi Data*) is a **2-credit, Diploma-level** course with **no prerequisites**. It's different from most statistics courses: there's very little calculation. Instead, you learn to **present data clearly and report it well** — a skill every employer now expects. By the end, you'll be able to build charts and dashboards in **Google Sheets and Power BI**, tell a story with data, and write and present a proper data report.

Three things you'll be able to do by the end (the course outcomes):

1. **Perform** simple plots and dashboards using data visualization tools effectively.
2. **Report and communicate** findings across different fields using effective visuals.
3. **Demonstrate** initiative and entrepreneurial skills in completing data-visualization tasks.

0.2 How you'll be assessed

There is **no final exam paper** in the usual sense — the course is 70% continuous work and a 30% practical final test.

Assessment	
Lab Test	
Group Project (report)	
Presentation	
Final Test (practical)	
Total	

> **Say it plainly to students:** "This course rewards *doing*, not memorising. You pass by building things — charts, a dashboard, a report, a presentation — not by writing essays in an exam hall. If you keep up with the weekly lab work, you're already most of the way to passing."

0.3 Ground rules, tools, and resources

- **Tools:** Google Sheets (free, browser — Weeks 5–7) and Power BI Desktop (free, Windows — Weeks 9–11). Ask students to install Power BI early.
- **Bring:** a laptop to lab sessions from Week 5 onward.
- **Note manual + slides** posted weekly; ebooks available (Dougherty, Wilke, Kusleika).

- **Attendance & lab habit:** the skills are cumulative — miss a lab and the next one is harder.
- Walk through the **14-week schedule** so they can see where the Lab Test (Wk 8), Final Test (Wk 13), and Project/Presentation (Wk 14) land, and start thinking about project groups early.

  Teaching tip. Keep the admin block to ~30 minutes. Students tune out of pure logistics — the fastest way to win them back is to move straight into the table-vs-chart demo below, which shows them *why* this course is worth their attention.

1.1 What is data visualization — and why does a statistician care about pictures?

Statistics is the science of making decisions when the data varies. We collect facts and figures, run procedures, and try to say something trustworthy about the world. But here's the catch: **a correct answer that nobody understands changes nothing.** A p-value on a slide that the audience can't read is a wasted analysis.

Data visualization is how we close that gap. It translates numbers — which live in the slow, analytical part of our brain — into shapes, positions, and colours, which our visual system reads almost instantly. A table of 500 rows is data. A chart that shows those 500 rows leaning in one direction is *understanding*.

> **Opening demo — 30 seconds, do this live.** Put a table of 12 monthly sales numbers on screen. Ask: "Which month was best?" Watch students scan the column. Then show the same 12 numbers as a line chart. The answer jumps out before you finish the question. That felt difference *is* the whole course.

Key idea: Visualization is not decoration added at the end. It is a *thinking tool* used throughout — first to help *us* understand the data, and later to help *others* understand our conclusion.

1.2 Two jobs, two mindsets: exploratory vs explanatory analysis

Every visual you ever make is doing one of two jobs. Getting students to name which job they're doing is one of the most useful habits this course can build.

Exploratory analysis — "a conversation between you and the data"

This is the messy, private, detective stage. You don't yet know what the story is, so you look from every angle: plot everything, sort, filter, check for outliers and gaps, test hunches. Most of what you make here is ugly and disposable — and that's fine, because the audience is *you*.

Think of it as sorting through 100 rough stones to find the one or two gems worth keeping. You make a *lot* of charts, quickly, and throw most away.

Explanatory analysis — "a conversation between the data and the audience"

Once you've found the gem, the job flips. Now you're not discovering — you're *communicating* one specific finding to one specific audience. You make *few* charts, but you polish them: clean labels, a clear title, everything unnecessary removed. The goal is no longer to learn; it's to make someone else understand and act.

Audience

Goal

How many charts

Mindset

Course chapters

> **Fresh worked example — Grab drivers in Kuala Lumpur.** A ride-hailing analyst has one year of trip data for KL. - *Exploratory:* She plots trips by hour, by day, by weather, by district — dozens of quick charts. Buried in them she notices demand spikes hard on rainy evenings in the city centre. - *Explanatory:* For the operations manager she makes **one** clean chart: average trips per hour on rainy vs dry evenings, with the rainy line clearly highlighted. Title: *"Rainy evenings drive 40% more trips in KL city centre — we need more drivers on standby."*
>
> Same dataset. First job: find the gem. Second job: show it.
>
> **Common misconception to correct.** Students hear "exploratory" and "explanatory" and assume you pick one. You don't — you almost always do **exploratory first, then explanatory.** Exploration finds the message; explanation delivers it. They are stages, not alternatives.
>
> Also flag the spelling trap: *explOratory* (explore) vs *explAnatory* (explain). They differ by one letter and students mix them up constantly on tests.

1.3 Before you draw anything: the three context questions

The single most common beginner mistake is opening a charting tool *first*. A good visualiser stops and answers three questions before touching a mouse:

1. **Who** are we communicating to? (One manager? A lecture hall? The general public? A minister?)
2. **What** do we want them to know or do? (One sentence. If you can't say it, you're not ready to chart.)
3. **How** can the data support that point? (Which numbers actually make the case?)

The audience decides *everything* downstream — how much detail, what tone, which chart, how much explanation. The same finding is drawn completely differently for a fellow statistician versus a Facebook post.

Fresh example — same finding, three audiences. Finding: *dengue cases in Selangor rose sharply in the third quarter.* - **To the State Health Director:** a detailed line chart, weekly cases, district breakdown, confidence bands. She wants precision to allocate resources. - **To the public on Instagram:** one big number — *"Dengue up 3× this month — clear stagnant water now"* — with a simple icon. They want a call to action, not a table. - **To fellow researchers:** the full time series with methodology notes and comparison to previous years.

One finding, three visuals, because three audiences. Ask students: *what changed between them?* (Answer: not the data — the purpose.)

In-class activity (Session 1 closer, ~15 min) — "Who's it for?" Give each group one finding (e.g. *"UiTM Raub students who attend tutorials score 12 marks higher on average"*). Assign each group a different audience: the Dean, first-year students, parents at open day. Each group sketches — on paper, not a computer — what their visual would emphasise. Groups present. The point lands hard: the audience, not the data, drives the design.

End of Session 1. Close by telling them: "Next session we get our hands on real data — how to meet a dataset, and the three reasons we bother visualising at all."

SESSION 2 — From Dataset to Evidence

Recap to open (~10 min). Ask two questions cold: *"What's the difference between exploratory and explanatory analysis?"* and *"What three things do you check before making any chart?"* Their answers tell you what stuck from Session 1.

1.4 Meeting a new dataset for the first time

Before visualization comes understanding *what you're holding*. When a dataset lands on your desk, resist the urge to chart it immediately. First, interrogate it:

- **What does it represent in the real world?** Where did it come from, and who collected it?
- **Who owns it,** and are you allowed to use and share it?
- **What do the rows and columns actually mean?** What period does it cover?

Then prepare the data before you explore:

- Check for **missing values** and **outliers**.
- **Rename** confusing variables; organise the data into a tidy shape.
- Compute **descriptive statistics** — measures of central tendency (mean, median, mode), position (quartiles, percentiles), variation (range, standard deviation), and skewness — to get a feel for the shape before you plot it.

Data attributes

An **attribute** is a characteristic that describes each record — usually a column. Naming attributes is a small step that prevents big confusion later.

Fresh example — a UiTM student dataset. Each **row** = one student. **Attributes (columns)** might be: *Student ID, Programme, Semester, State of origin, CGPA, Tutorial attendance %.* For a **kopitiam's daily sales dataset**, each row = one day, and attributes might be: *Date, Day of week, Total sales (RM), Number of customers, Weather, Public holiday (Y/N).*

Ask: *before charting, what would you check?* (Missing sales on some days? A CGPA of 4.5 that can't exist? A "state" typed three different ways?) This is where good visualization actually begins.

Live walkthrough (~20 min). Open a real spreadsheet on screen (use the Week 1 dataset in the folder). Walk the class through it out loud: name the rows and columns, spot a missing value, point out an outlier, read off the mean and median. Frame visualization as *seeing* the descriptive statistics they already compute — a boxplot is a five-number summary you can look at; a histogram is a frequency table you can see. This connects straight back to their prior stats knowledge.

1.5 What the experts say data visualization is

It helps students to hear that this isn't just your opinion — the field's leading authors converge on the same idea: **visualization is storytelling with data.**

- It **summarises complex raw data into a simpler, understandable form.**
- It is **a graphical representation of data or concepts** (Ware, 2008).
- When a graph is presented well, **information flows most clearly and efficiently to the viewer** (Wong, 2010).
- It is **a narrative that answers a question without unnecessary detail** (Fry, 2008).
- **The numbers are the source material; the diagram is how we describe them** (Yau, 2013).

Pull the common thread out loud: every one of these says *clarity in service of a message.* Not "make it pretty" — make it *clear*.

1.6 Why we visualize: three reasons

Give students a simple three-part answer they can recall on a test: we visualize **to communicate, to transform data into information, and to show evidence.**

1. To communicate

A table is good for looking up *one* exact value. But the moment you want to see a *pattern* — a trend, a relationship, an oddity — you need shape, not numbers. How we encode data (dots, lines, bars, areas, bubbles) reveals things that stay hidden in a table:

- **Dots** in a scatterplot → a relationship (more revision hours → higher marks).
- **Lines** in a line chart → a trend over time (monthly sales climbing).

- **Bars** in a bar chart → a comparison (which branch sells most).
- **Shaded areas / bubbles** → clusters or anomalies (one freak day of huge sales).

2. To transform data into information

In the age of Big Data, raw numbers are overwhelming — Big Data is not just *a lot* of data, it's data too large and fast to grasp unaided. Visualization is how we move up the ladder: **data → information → knowledge → action.** A chart that reveals an insight, and then drives a decision, has done real work.

> **Fresh example — the ladder in action.** - **Data:** 12 months of a clinic's patient counts. - **Information:** a line chart shows visits spike every year in the haze season. - **Knowledge:** respiratory cases drive the spike. - **Action:** the clinic pre-orders inhalers and adds staff before the haze arrives.
>
> The chart is the rung that turns numbers into a decision.

3. To show evidence

A data graphic is *visual evidence*. When you claim "sales are recovering," a clean upward line is proof the audience can see for themselves. But evidence is only as good as the insight behind it — a beautiful chart of a meaningless number persuades no one who's paying attention. **A chart is only as good as the message it carries.**

Chapter summary (student-facing recap)

- Visualization bridges the gap between a correct analysis and an audience that understands it.
- **Exploratory** analysis (you + data, find the gem) comes first; **explanatory** analysis (data + audience, show the gem) comes second.
- Before charting, always answer: **who** is the audience, **what** is the message, **how** does the data support it.
- Meet a dataset properly first: know its source, clean it, compute descriptive stats, name its attributes.
- We visualize for three reasons: **to communicate, to transform data into information, and to show evidence.**

Key terms

Exploratory data analysis (EDA) · **Explanatory data analysis** · **Context (audience / message / mechanism)** · **Attribute** · **Descriptive statistics** · **Big Data** · **Data → information → knowledge → action**

Discussion questions

1. Give an example from your own life (a WhatsApp group, a game, a part-time job) where you explored data before explaining it to someone.
2. Why might the *same* finding need two completely different charts?

3. A friend says "I'll just make the chart first and figure out the point later." What would you tell them, and why?
4. Where does data visualization sit on the ladder from data to action?

End of Session 2 — preview Chapter 2: "Now that you know *why*, next week we learn *which chart* to use — and the traps that make good data look bad."

Timing at a glance

Session 1 (~110 min): Course briefing 30 · §1.1 hook & what-is-viz 15 · §1.2 exploratory vs explanatory 25 · §1.3 three context questions 15 · "Who's it for?" activity 15 · buffer/wrap 10.

Session 2 (~110 min): Recap 10 · §1.4 meeting a dataset + live walkthrough 35 · §1.5 scholarly perspectives 10 · §1.6 three reasons to visualize 25 · discussion questions 15 · recap + Ch 2 preview 15.

Draft notes for Suriyati: figures referenced in the original manual (Pareto "computer problems", Figures A–D) are not reproduced here — flag me where you want new figures and I'll spec them or build them. All worked examples above are new (Grab-KL, Selangor dengue, UiTM students, kopitiam sales, haze clinic) and do not reuse the manual's examples. Assessment figures are from the CAP/SOW; confirm the exact Lab Test/Final Test split matches your section's plan.

Glossary

- **Exploratory data analysis** — looking at data from many angles to discover what's interesting, for yourself.
- **Explanatory data analysis** — communicating one specific finding to a specific audience.
- **Context** — the three questions before charting: who, what, how.
- **Attribute** — a characteristic describing each record, usually a column.
- **Descriptive statistics** — mean, median, quartiles, spread, skewness.
- **Big Data** — data too large or fast to grasp unaided.
- **Data → information → knowledge → action** — the ladder a good chart climbs.