EXAM REVISION SHEET

Full guide + ask Sia about ETF5922

ETF5922 Data visualisation and communication

MONASH UNIVERSITY · SEMESTER 2, 2026

SIDE 1 · RECALL + DECIDE

Final exam 50% · individual examination

METHODS BEFORE PIXELS

0 · Exam map

COURSE.JSON · BRIEF

Q 10 | A1 10 | A2 15 | A3 15 | FINAL 50%

Quizzes 10% · Assignment 1 10% · Assignment 2 15% · Assignment 3 15% · Final exam 50%.

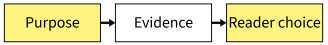
OFFICIAL-RULE STATUS · BRIEF

```
Permitted materials: not confirmed
for this offering in the course
materials we hold — check the exam
instructions on the LMS.
```

1 · Purpose → claim

CH 1

★ *Exploration* is rapid, provisional plotting to discover structure. *Communication* selects evidence and context so another reader can recover the intended comparison without narration.



Source figure, CH 1: purpose fixes the evidence and comparison.

MODE	KEEP / CHANGE
Explore	many candidate views; rough labels acceptable
Communicate	stable labels, selected evidence, context and limitation
Across audiences	keep values, units, order and calculation base; change hierarchy and detail

ANSWER SCAFFOLD

```
Reader + decision → visible
comparison → measured support →
boundary of claim.
```

1b · Annotation audit

CH 1 FAQ

Hide the annotation and ask a fresh reader for the main comparison; restore it and test whether accuracy or speed improves. Keep labels that identify evidence, context or uncertainty; remove repetition that competes with the focal mark.

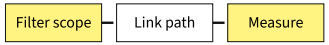
BRIEF FIELD	WHAT EARNS ITS PLACE
reader	prior knowledge and decision conditions
decision	comparison/action the graphic must support
scope	observations, time window, unit and denominator
evidence	values and context sufficient for the claim
limit	main uncertainty or explanation outside scope

Release test: a communication graphic must survive separation from its maker. Rebuild when question, comparison base or limitation cannot be recovered unaided. When several audiences share a page, keep the common numerical core and reveal extra detail through hierarchy or optional notes.

2 · Power BI context

CH 2

★ *Filter context* is the complete set of report, page and visual filters governing a displayed calculation. A **measure** recalculates under that context; a stored column retains row-level values.



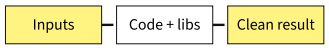
Source figure, CH 2: every card inherits a filter path.

AUDIT ORDER	QUESTION
1 · source rows	what total should these rows produce?
2 · keys	unique, duplicated or unmatched?
3 · relationships	which direction propagates selection?
4 · state	which slicers/filters are active?
5 · output	do card, table, chart and title reconcile?

Unexpected totals can arise from duplicate keys, bidirectional paths or implicit aggregation even when the page looks polished. A static export must print the selected subset or date range because controls may disappear.

3 · Reproducible R

CH 3



Source figure, CH 3: clean-session provenance.

Atomic vector	one dimension; elements share an underlying type
NA	unavailable, not zero or an empty string; declare treatment
Factor	levels control categories/order; ordered factors permit ordinal comparison
Project root	stable relative paths, not personal machine paths

COMPLETENESS BESIDE CENTRE

```
mean = Σ observed values / n
observed
missing % = n missing / n expected ×
100
```

Restart and run from the top. A vanished object exposes hidden console state; a second clean environment must restore dependencies, locate inputs and reproduce outputs.

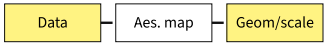
Availability is part of the result: a mean without the observed count cannot show whether expected records contributed. Report the calculation and its missing-value condition together. Keep factor levels deliberate: character-to-factor conversion may change order, summaries and plot structure.

CONTRAST *Power BI requires visible interactive context; R requires declared computational context. In both, the result must be reconstructable away from the analyst's current screen.*

4 · Grammar of graphics

CH 4

A plot is an explicit construction: **data** supply observations/variables; **aesthetic mappings** connect variables to visible channels; **geoms** draw marks; scales, facets, coordinates and themes complete the view.



Source figure, CH 4: the layered visual sentence.

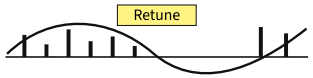
MAPPING VS SETTING

```
inside aes(…) → a variable controls
the channel
inside geom(…) → every mark receives
the same appearance
```

READER TASK / VARIABLE	DEFENSIBLE DISPLAY
category magnitude	bar / dot; aligned length or position
continuous distribution	histogram + rug; density + raw support; boxplot for landmarks
two quantitative variables	scatter; position carries both
ordered time	line/points with interval and gaps visible
groups with common comparison	small multiples on fixed scales

4b · Distribution sensitivity

CH 4 FAQ



Source figure, CH 4: compare smoothing with raw locations.

Bin width changes which observations are pooled; **bandwidth** changes density smoothing. A mode, gap or cluster is defensible only if it survives nearby reasonable settings and remains visible in raw support.

DISPLAY	CARRIES / COMPRESSES
rug	exact locations / little summary
histogram	pooled counts / depends on boundaries and width
density	smooth shape / bandwidth can create or conceal peaks
boxplot	centre, spread, potential outliers / hides local clusters

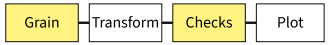
Decision rule: choose the display whose perceptual task matches the question, then add a complementary view when a transformation or smoothing choice could hide the observations supporting the claim.

TRAP *Defaults are choices. Report a tuning-sensitive impression cautiously; never delete raw evidence to stabilise the story.*

5 · Wrangling = argument

CH 5

★ *Observation grain* is the entity represented by one row at the current stage. State it before transforming: the chart inherits every population, denominator and duplication error in its table.



Source figure, CH 5: audit the table before the chart.

VERB	EXPECTED STRUCTURAL EFFECT
select	columns change; rows preserved
filter	declared rows removed; population changes
mutate	variables added/changed; rows preserved
group + summarise	many source rows → one row per group; carry observed/missing counts
pivot	wide ⇔ long layout; substantive measurements unchanged
join	row count follows key cardinality; unmatched keys may disappear

LOCALISE AN ERROR

```
record n rows + key uniqueness after
each verb; the first unexpected
change identifies the operation to
inspect.
```

5b · Join gate

CH 5 FAQ

BEFORE	AFTER
name both keys; test uniqueness; predict 1:1, 1:M or M:M; record expected n	inspect unmatched and multiplied rows; reconcile important totals; confirm grain

A left join grows when a right-side key has multiple matches. Do not repair the plot while leaving duplicated records in the data.

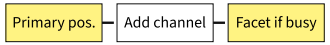
CARDINALITY	ROW-COUNT IMPLICATION
1:1	at most one match on either side; retained keys remain single rows
1:M	the "one" record repeats once per matching "many" row
M:M	combinations multiply; require an explicit substantive justification
unmatched	retained or discarded according to join type; count explicitly

Summarise gate: grouping changes row meaning to one row per group. Carry `n observed` and `n missing` into the output so a clean bar does not conceal unequal evidence support.

EXAM MOVE *Explain what one row means before and after summarising or joining; then say how the changed denominator affects the visual claim.*

6 · Encode more variables

CH 6



Source figure, CH 6: add a channel only for a named question.

CHANNEL	USE / WARNING
position	most precise quantitative comparison; reserve for primary variables
length	good magnitude comparison on a common baseline
hue / shape	category identity; test greyscale and output size
lightness	ordered magnitude when the palette is sequential
size / area	coarser judgement; avoid if exact comparison matters
labels	lookup; direct labels can remove legend travel

Palettes: qualitative for categories; sequential for ordered low→high; diverging around a meaningful midpoint. An accessible palette preserves the required comparison under common colour-vision differences and realistic reproduction.

6b · Facet decision

CH 6 FAQ

Fixed scales	cross-panel magnitude is the task; common ruler retained
Free scales	local shape is the task; warn that panel magnitudes are not directly comparable
Caption	name each subset, scale choice, comparison and channel meaning

If removing a channel loses no variable, remove it. Facets reduce overload only when repeated meanings and labels remain consistent.

7 · Perceptual honesty

CH 7

Data ink carries values or essential guides; **chartjunk** competes without improving interpretation. Data-to-ink reasoning is not indiscriminate minimalism: keep context needed for accurate reading.

LIE FACTOR

```
apparent proportional change /
actual data proportional change
target: 1; state whether geometry
uses length, area or another
dimension.
```

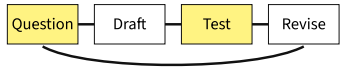
PERCEPTUAL RULE	EXAM CONSEQUENCE
position/length beat angle/area/hue	assign precise quantities to the most accurate channel
preattentive contrast directs search	reserve emphasis for the focal evidence
proximity, similarity, enclosure, connection group marks	make perceived groups match data structure

8 · Pattern + support

CH 8–9

Explore distribution, association, clusters, unusual observations and gaps together. A centre without spread hides heterogeneity; a smooth without raw points hides support; a mean without completeness can imply false precision.

IF YOU SEE…	DO THIS
outlier	retain while checking measurement, context and influence; report sensitivity
small/unequal groups	show raw points, observed n and missing proportion
association	describe co-variation; do not upgrade it to cause
interval	name what variability/support it represents and its assumptions
unstable conclusion	vary one defensible treatment and report what persists



Source figure, CH 8: controlled revision needs a reading test.

Analytical critique connects a visible choice to a consequence for accuracy, comparison or readability, then proposes one targeted repair. Stop when the intended reader recovers the comparison, values/scope reconcile, and every remaining element has a role.

9 · Time + interaction

CH 10–11



Source figure, CH 10: no observation means no connecting line.

Time	state frequency, window, gaps, aggregation and baseline before trend/seasonality
Indexing	for relative growth from a common start; label base and retain raw units
Interaction	expose filter/selection/zoom state and provide a documented reset
Tooltip	optional detail only; units, scope and key evidence remain visible
Animation	stable axes, identities and reference points; controls plus static fallback
Small multiples	prefer for precise comparisons across several moments

THREE-SENTENCE INTERPRETATION

```
1 pattern visible · 2 quantified
support · 3 uncertainty/limitation.
```

FINAL CHECK *A short sequence may support endpoint changes, not annual seasonality; an event annotation locates coincidence, not impact.*

EXAM REVISION SHEET

Full guide + ask Sia about ETF5922

ETF5922 Self-test drills · answer paths · classic traps

DATA VISUALISATION AND COMMUNICATION · MONASH UNIVERSITY

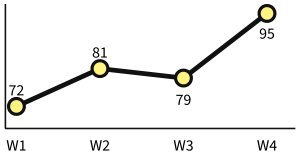
SIDE 2 · WORK + CHECK

Five drills · method chooser · exam-morning plan

SHOW DENOMINATOR + LIMITATION

D1 · Measured claim CH 1 WORKED

Data: weekly service completions `72, 81, 79, 95`. The director wants the latest change visible.



CH 1 numbers: labelled line preserves observation order.

1. **Latest absolute:** `95 - 79 = 16 cases`.
2. **Latest relative:** `16 / 79 = 20.3%`; `79` is the stated base.
3. **First→last:** `95 - 72 = 23`; `23 / 72 = 31.9%`.
4. **Encode:** labelled line with points retains all four weeks.
5. **Qualify:** observed change alone does not identify a cause.

ANSWER

```
Latest completions rose by 16 cases (20.3% from 79); W4 is 23 cases (31.9%) above W1. The sequence describes change, not cause.
```

SELF-CHECK *Every percentage names its denominator; every causal-sounding verb is removed.*

Absolute ≠ relative CH 1 · 11

absolute change	new − old; retain original unit
relative change	(new − old) / old × 100%
percentage points	new % − old %

For a share `30% → 45%`: change = **15 percentage points**; relative change = `15/30 = 50%`. Do not write "15% increase".

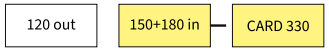
Claim quality grid CH 1 · 9 · 10

CHECK	PASS CONDITION
question	named reader and comparison; chart answers that task
quantity	number, unit and denominator independently recomputed
scope	time window, group and observation grain visible
support	raw points/counts/completeness available where summaries could mislead
language	description remains proportionate to design and evidence
limit	main uncertainty, missingness or causal boundary stated

Cold-read test: cover the subtitle, hand the view to a fresh reader and ask for the main comparison, base and limitation. If oral rescue is needed, revise labels, ordering or context—not the recorded values.

D2 · Filtered card CH 2 WORKED

Data: regional sales `120, 150, 180`; slicer retains the final two regions.



Source figure, CH 2: visible filter scope produces 330.

1. Resolve scope before adding: retained values are 150 and 180.
2. `Filtered = 150 + 180 = 330`.
3. `Full = 120 + 150 + 180 = 450`.
4. `Retained share = 330 / 450 = 73.3%`.
5. Place the two-region state beside the card; title, table and chart must agree.

ANSWER

```
330 under the active two-region context; 450 overall; 73.3% retained. The displayed total is conditional on the filter.
```

FAILURE	INSPECTION
wrong card	rows → aggregation → filters → relationship path
different screenshot	slicer/date state absent from export
inflated total	duplicate key or many-to-many match

Power BI judgement CH 2 FAQ

If a result must change with report context → use a measure. **If it is a row-by-row property →** a stored column may be appropriate. The analytical role decides, not interface convenience.

STATE	EVIDENCE TO EXPOSE
opening	whether all data or a curated subset is shown
selection	active regions, dates or categories beside headline values
reset	clear return to the documented starting state; title and annotations reset too
export	scope survives when slicers or hover controls vanish

Reconciliation proof: manually write contributing records and expected sum; clear/reapply the selection; verify coordinated cards, tables, charts and titles. Validation requires both the number and the path that produced it.

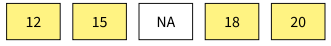
TRAP *A visible slicer is evidence, not decoration: without it, neither reader can reconstruct why the card changed.*

Dashboard-to-data trace CH 2 · 5

1. State what one source row represents and which field is the grouping key.
2. Predict row and total effects before filtering, grouping or joining.

D3 · NA + clean session CH 3 WORKED

Vector: `12, 15, NA, 18, 20`. Report centre and completeness.



Source figure, CH 3: NA is unavailable, never zero.

1. Observed count `n = 4`; expected count 5.
2. Observed sum `12+15+18+20 = 65`.
3. Mean `65 / 4 = 16.25`.
4. Missingness `1 / 5 = 20%`; completeness = 80%.
5. Restart the project and reproduce both values without prior console objects.

ANSWER

```
Mean 16.25 from four observed entries; one of five values missing (20%). Clean-session rerun establishes declared provenance.
```

Clean-session checklist CH 3 FAQ

inputs	declared, stable, separate from derived outputs
packages	recorded and loaded by the script
objects	created in execution order
paths	project-relative, not personal
missingness	explicit policy plus observed/expected n
portability	second clean environment reproduces result

Factor trap: category levels and their ordering can change summaries and plots. Conversion is an analytical decision, not harmless formatting.

SELF-CHECK *A copied workspace hides dependencies; reconstruction from data + code demonstrates reproducibility.*

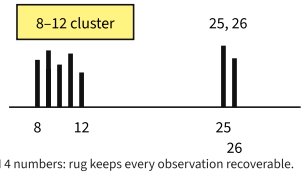
Wrangling under pressure CH 5

OPERATION	QUESTION BEFORE MOVING ON
filter	which population and denominator were removed?
mutate	does the new variable preserve row count and units?
group	which keys define each resulting comparison?
summarise	are observed n and missing n carried forward?
pivot	did layout change without changing substantive measurements?
join	are keys unique; what cardinality and unmatched count are expected?

For sales `40, 60, 80, NA` in regions A, A, B, B: A totals 100 from two rows; B totals 80 from one observed row and one missing row, so B is 50% complete. The plot must not present 100 versus 80 as equally supported.

D4 · Distribution tuning CH 4 WORKED

Times: `8, 9, 9, 10, 11, 12, 25, 26 minutes`.



CH 4 numbers: rug keeps every observation recoverable.

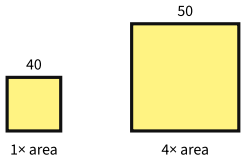
1. List raw observations before selecting the distribution display.
2. `Σx = 110; n = 8; mean = 110/8 = 13.75 min`.
3. Note the 8–12 cluster and separated high pair 25, 26.
4. Pair histogram/density with rug or another raw layer.
5. Vary bin width/bandwidth; call a second mode only if it survives defensible tuning.

ANSWER

```
Mean 13.75 min is pulled upward by 25 and 26. Raw locations must accompany smoothing; any claimed mode is a sensitivity result.
```

D5 · Geometric distortion CH 7 WORKED

Data: `40 → 50`; pictogram doubles height and width.



CH 7 numbers: two doubled dimensions quadruple area.

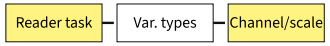
1. Data change `(50-40)/40 = 25%`.
2. Area factor `2 × 2 = 4`; apparent area increase = 300%.
3. Lie factor `300% / 25% = 12`.
4. Repair with one-dimensional aligned length proportional to values.

ANSWER

```
The geometry suggests +300% while data rise +25%; aligned length restores proportional comparison.
```

Chooser · If X → do Y CH 4–11

PROMPT SIGNAL	FIRST DEFENSIBLE MOVE
category ranking	sort by the task; aligned position/length; direct labels
continuous distribution	raw support + histogram/density/boxplot; tune deliberately
association	scatter; show raw points; qualify causation
third category	accessible hue/shape; remove unused channels
many groups	small multiples; fixed scales for magnitude
unequal missingness	observed n + missing % beside summaries
ordered time	chronology + interval + gaps; common scale
relative growth	index to declared base; retain raw units
transition itself matters	animation with stable references + controls + still alternative
exact cross-frame compare	aligned static frames, not motion memory
dashboard total	reconcile rows, keys, relationships, filters and aggregation
chart criticism	choice → reader consequence → one repair → reading test



Source synthesis: task and types choose the encoding.

Small-multiple test CH 6

Three means `12, 18, 21` give a nine-unit range. Position carries the means; accessible hue or shape can carry status. If faceted for cross-group magnitude, keep a shared scale.

SAY WHY

```
"Fixed scales preserve the common ruler needed to compare 12, 18 and 21; a free scale would switch the task to within-panel shape."
```

TRAP *More encodings are not more information. Add a channel only when removing it would lose a necessary variable.*

Uncertainty judgement cards CH 9

If a mean looks stable but n is small →	show raw points, observed n and completeness; avoid implied precision
If one point is far away →	retain while checking measurement/context; compare influence with it flagged
If groups have different gaps →	put missing proportions beside summaries before reading the contrast

Classic traps ALL CHAPTERS

TRAP	REPAIR SENTENCE
pattern → cause	"The display describes association/change; the evidence does not identify a cause."
hidden denominator	state base and distinguish points from relative %
hidden filter	print active subset/date beside the headline
NA treated as zero	report observed n, expected n and missing policy
default smoothing	retune and retain raw marks
join inflation	test keys/cardinality; reconcile n and totals
free facets read as equal	fix scales or warn against magnitude comparison
area implies value	compute distortion; repair with aligned length
outlier deleted	retain, investigate, report sensitivity
four months = seasonality	state interval/window and comparable cycles

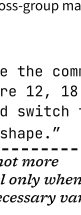
Time-series guardrails CH 10

Order and interval are data. Aggregation can suppress peaks; indexing replaces magnitude with relative growth; moving summaries reduce noise while concealing short events. Preserve missing periods as breaks. If interpolation is justified, distinguish estimated from observed values and name the method.

BEFORE NAMING TREND/SEASONALITY

`frequency · window · gaps · aggregation · baseline · number of comparable cycles.`

Exam-morning plan COURSE STRATEGY



1. Question + grain
2. Wrangle + audit
3. Encode comparison
4. Recompute evidence
5. Pattern + support + limit

Answer path from the course revision strategy.

1. Underline question, variable types, unit, scope and denominator.
2. State one-row grain; predict row effects; check keys and missingness.
3. Choose mark, channel, scale and order; show active state.
4. Recompute one anchor; verify labels, totals and base.
5. Write pattern; quantified support; uncertainty or boundary.