

Week 6: Roadmap, Planning Rituals, and the Data Story

DATA 510: Data Science Capstone

Lucas P. Cordova, Ph.D.

2026-06-15

Week 6, in the gap between the approved proposal (M1) and the Data Summary (M2). A 60-minute briefing in three parts. The Roadmap: where each milestone becomes a section of the final write-up, reviewed by me and your peer POs, and what the Data Summary requires. Planning Tonight: how iterations are sized to each milestone, how studios run at milestone boundaries, and how to run tonight's first 20-minute planning session on the board and in the repo. The Data Story: told as a user story and built on a write-up skeleton you grow milestone by milestone.

Table of contents

1 Learning Objectives	1
2 Part 1: The Roadmap	2
3 Part 2: Planning Tonight	7
4 Part 3: The Data Story	13
5 Wrap	15

1 Learning Objectives

1.1 Today's Objectives

1.2 What You Will Leave With

By the end of this hour, you will be able to:

1. **Place** your project on the roadmap and explain how each milestone becomes a section of the final write-up.
2. **State** what the Data Summary (M2) must demonstrate, and how iterations are now sized to each milestone.
3. **Plan** tonight's iteration: Ready cards on the board, each with a Create / Observe / Analyze triple, a milestone tag, and a size.
4. **Frame** your project as a data story, told as a user story, and place it in a write-up skeleton.

1.3 Course Connection

Everything you produce is reviewed by **me and your peer Stakeholder POs**, and every milestone is a section you bank toward the **final write-up**. This hour sets up your **Data Summary** (M2) push and the studio rhythm you will run at every milestone boundary. It carries Course Objective 1 (development practices for publication) and Course Objective 4 (communicating findings clearly).

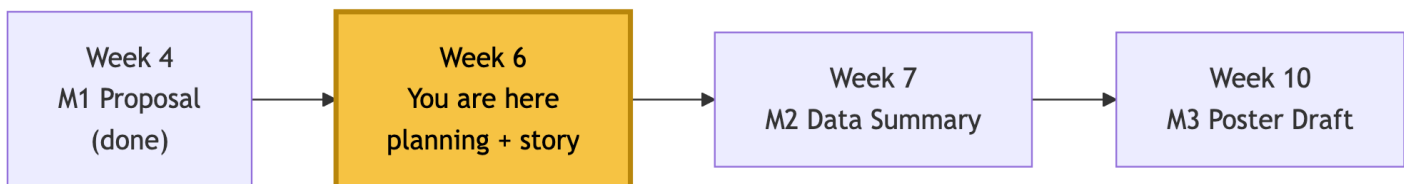
2 Part 1: The Roadmap

The proposal was the promise. Here is how we keep it.

2.1 The Roadmap

2.2 You Are Here

The proposal is in. The Data Summary is next, at the end of week 7. Two class meetings stand between you and that gate.



The proposal told me what you intend to do. Everything from here is you doing it, one milestone at a time.

2.3 Every Milestone Is a Section of the Write-Up

Nothing you submit this term is throwaway. Each milestone is a **section you bank** toward the final write-up. The proposal is already your first section: the project description, the motivation, the question, the stakeholders.

M1 Proposal

intro + moti

M2 Data Summary

data + engin

M3 Poster Draft

results + s

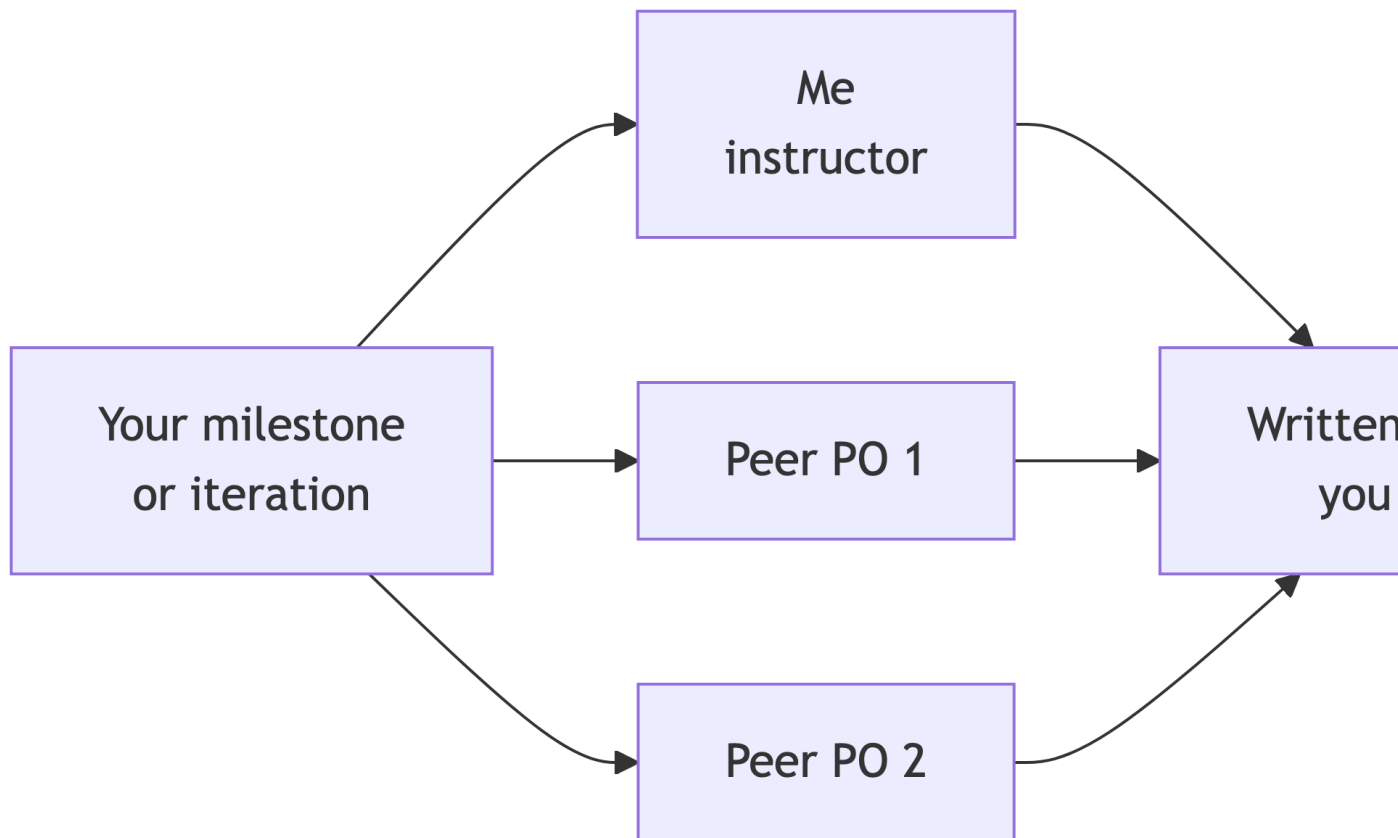
M4 Write-up Draft

full assem

By M4 you are assembling and polishing sections you already drafted, not staring at a blank page.

2.4 Who Reviews Your Work

One thing worth stating clearly before we plan: **nothing you produce is reviewed in private by me alone.**



Everything I review, **your peer Stakeholder POs review too.** The people critiquing you are the same people who will read your write-up sections as they accrete. Write every section so a peer can navigate it.

2.5 Proposal Feedback Is in Flight

A timing note: **I am reading your proposals this week.** My written feedback is coming back to you while you are already planning toward the Data Summary.

...

Do not wait for it before you start. This is an iterative process, and the proposal is your default fallback, not a fixed contract. **Anything I flag can be folded into the Data Summary** (M2): adjust scope, document the delta, and keep moving.

2.6 What the Data Summary Requires

The M2 deliverable is one PDF in `deliverables/M2-data-summary/` that points to repo paths. Seven sections.

#	Section	What it proves
1	Research question reminder	The data scope still answers the proposal's question
2	Data inventory	Source, license, size, date range, refresh policy, repo location
3	Schema and organization	Keys, join paths, raw vs analysis-ready layers
4	Ingestion status	Manual pulls done, or a pipeline that runs without babysitting
5	Data quality and profiling	Missingness, duplicates, outliers, and what you did about them
6	Reproducibility	Commands to rebuild analysis-ready data; environment notes
7	Ethics and access controls	PII, consent, redaction, sharing constraints

This is section 5 of your write-up (“Data and data engineering”) being born early. After M2, the work shifts from acquisition to analysis.

2.7 Iterations Are Sized to the Milestone

Here is a change from a standard weekly cadence: an **iteration is sized to the milestone it serves**, not to a fixed week.

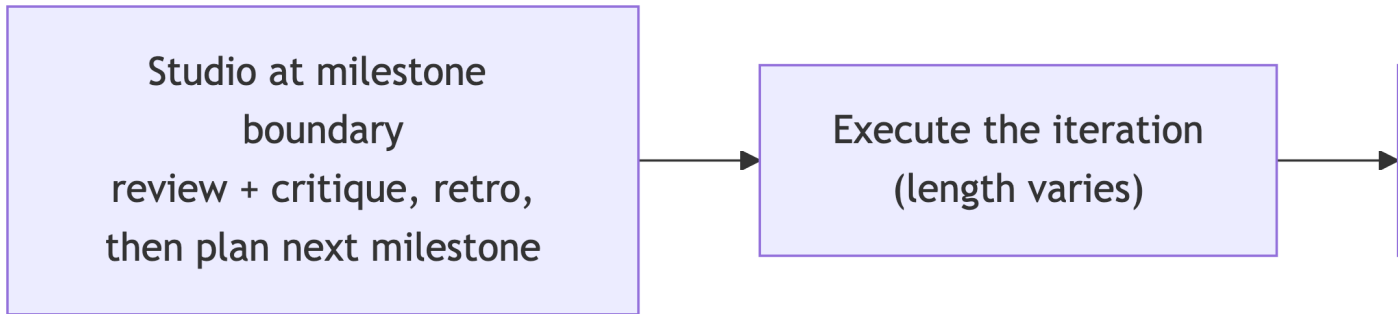
- A tight milestone like the Data Summary gets a **short iteration**.
- A long stretch like the run to the poster gets a **longer iteration**.

...

You plan the iteration to fit the milestone in front of you. The board and the Create / Observe / Analyze discipline stay the same; only the length changes.

2.8 Studios Run at Milestone Boundaries

Because iterations are milestone-sized, **studios happen at milestone boundaries**, not every single week. A single studio does **all three rituals**: review and critique, retrospective, then plan the next milestone.



Think of each studio as a **bookend**: it closes the milestone just finished and opens the next one. **Tonight is the exception**: it is the first studio, so there is nothing to review or retro yet, just planning for the Data Summary. The next studio is the **week 8 class**, after the Data Summary is in: there you review, critique, retro, and plan what comes next.

2.9 Quick Quiz: What “Done” Means for M2

Which project has met the Data Summary bar?

- A. "We found three great datasets and bookmarked them."
- B. "Our scraper runs nightly, lands clean tables in data/processed/, and we documented the s
- C. "We have all the raw files downloaded, about 40 of them, unsorted."
- D. "We trained a first model and it looks promising."

...

B. M2 is about the data being **in, organized, and self-sustaining**, with enough documentation that I trust it. A is sources, not data. C is acquisition without organization. D jumps past the gate.

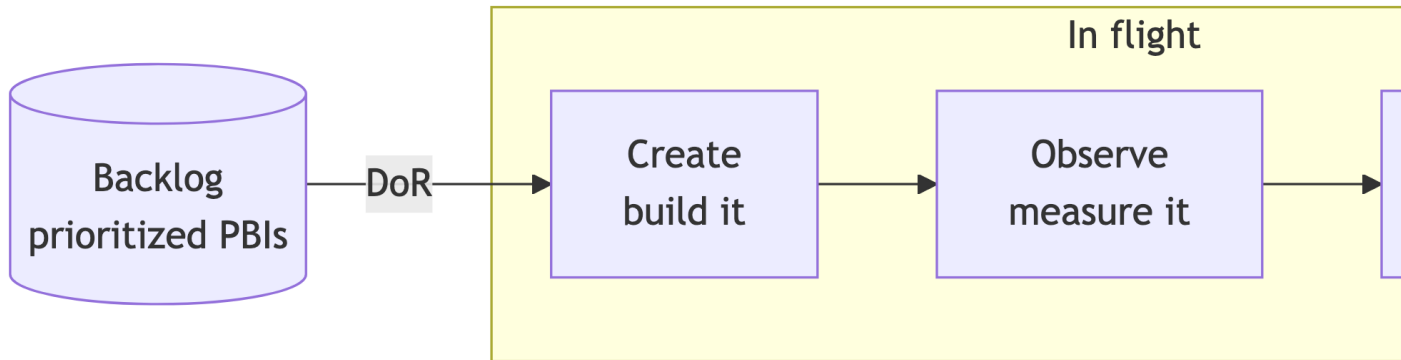
3 Part 2: Planning Tonight

A plan you do not write down is a wish. Here is the ritual that turns wishes into cards.

3.1 Planning Tonight

3.2 The Board Is Where Planning Lands

Every card moves across the same five columns. Planning means deciding which cards earn a spot in flight.



The columns are not job types. They are **phases of one card**: build it, measure it, decide what it means.

3.3 The Create / Observe / Analyze Triple

The triple is for **data PBIs**: cards that answer a question about your data. For the Data Summary, that means **acquisition and profiling** work.

Phase	What you produce (M2 flavor)
Create	The artifact: an ingestion script, a cleaned table. <i>Something exists in the repo.</i>
Observe	Evidence about it: row counts, null rates, coverage, duplicates. <i>A measurement is recorded.</i>
Analyze	The data-quality decision: impute, flag, drop, re-pull, or accept. <i>A next step is named.</i>

...

For M2, “Analyze” is **not** modeling. It is deciding what the profiling tells you to do about the data. Modeling and statistics come at M3 and later.

3.4 Not Every Card Is a Data PBI

Some cards are plain **setup or infra**: create the repo, add a `.gitignore`, invite peer POs to the board, wire up CI.

...

These have no meaningful Observe or Analyze, so do **not** force a triple onto them. Tag them **infra**, keep them on the board for visibility, and treat them as simple **done / not-done**.

Card type	On the board	Needs a triple?
Data PBI (acquire, clean, profile)	Backlog -> Create -> Observe -> Analyze -> Done	Yes
Infra / task (setup, glue, admin)	Backlog -> Create -> Done	No, done / not-done

This iteration, most of your meaningful cards are **data PBIs** (acquisition + profiling), with a few **infra** cards to stand up the plumbing.

3.5 The Definition of Ready

A card should clear this bar before it moves into **Create**:

- A one-sentence hypothesis or user story (**data PBIs**)
- A named **Create / Observe / Analyze** triple (**data PBIs**; infra cards skip this)
- A milestone tag (this iteration, almost always **M2-data-summary**, or **infra**)
- A size (S, M, L, XL)

...

Infra cards still need a tag and a size, just not a hypothesis or a triple. For a data PBI, if it cannot answer all of these, it stays in **Backlog** and we refine it.

3.6 Tags and Sizing

Milestone tags

Every card carries **M1..M5**, **infra**, or **ethics**. This iteration, lead with **M2-data-summary**.

Sizing for a milestone iteration

Size each card S to XL, then pull only what fits the **length of this milestone's iteration**. The Data Summary runs **two weeks** (due end of week 7), so size cards to finish within that window.

3.7 Tonight: Your First Studio Session

This is the **first** studio of the term, so we keep it simple: **planning only, 20 minutes per project**. No critique tonight, because you already did that in your peer reviews of the proposal. The goal is one committed iteration plan aimed at the Data Summary.

. . .

This briefing is the setup. After it, each project gets its 20 minutes to plan with its peer POs at the table.

3.8 What Each Role Does Tonight

Tonight everything happens **in the 20-minute meeting**. Owners build the plan live; peer stakeholders give input live and file their written review afterward.

Owner team (in the meeting)

- Open your **board and backlog** from the charter and draft a few candidate M2 cards.
- Frame what your **Data Summary needs** next.
- Drive the session and commit the plan.

Peer stakeholders (live, then review)

- Nothing beforehand.
- Give live input on the iteration: requirements, questions, risks; help prioritize.
- Commit your written review to the repo (**studio/**) afterward.

3.9 The 20-Minute Agenda

Phase	Time	Who leads
Owner framing: project + M2 goal + candidate cards	5 min	Owner team
Peer PO input on the iteration	6 min	Peer POs
Backlog refinement and board walk	7 min	Owner team, with peer PO input
Commit the iteration plan	2 min	Owner team

The planning happens in the board walk: refine the top of **Backlog** to Ready, then pull the cards you can realistically finish this iteration. You leave with a committed plan.

3.10 Example: A Studio in Action

Let me walk one studio with three borrowed faces. The names are for fun; the moves are the point.

- **Sheldon**, Owner Product Lead. His project forecasts weekly **Pasadena bus ridership**.
- **Penny**, peer stakeholder. She presses on *who decides what*.
- **Howard**, peer stakeholder. He presses on *the data and the pipeline*.

. . .

Tonight is **only iteration planning for M2**. There is no retrospective yet; we run the retrospective ritual before the next studio, once there is an iteration to look back on.

3.11 Example: Owner Framing and Peer Input

Sheldon (owner, 5 min): “My project forecasts weekly ridership for Pasadena Transit. For the Data Summary I need every 2020 to 2024 ridership record pulled, profiled, and documented. Three candidate cards: pull the API, profile the nulls, document the schema.”

. . .

Penny (peer, input): “Who actually uses this forecast, and to decide what? Say it as a user story.”

. . .

Sheldon: “As a Pasadena Transit planner, I need to know next month’s ridership by route, so that I can set bus frequency.”

. . .

Howard (peer, input): “Does the API really cover 2020? Ridership cratered that year. If there are gaps, you decide impute versus flag *before* you call the data done, not after.”

3.12 Example: Board Walk and Commit

Sheldon (board walk, 7 min): refines the top card to Ready.

- 1 Hypothesis: The Pasadena Transit API covers 2020-2024 well enough to skip manual export.
- 2 Create: Script pulling all 2020-2024 records into data/raw/ridership/.
- 3 Observe: Records per year vs the agency's published totals; flag any year off by >5%.
- 4 Analyze: Decide impute vs flag for low-coverage years.
- 5 Tag: M2-data-summary Size: M

...

He is solo, so he keeps it small: he pulls the **API card** and the **profiling card** (both data PBIs) into **Create**, and **defers** the schema-doc card so he is not juggling too much at once.

...

Howard: “What about the data/processed/ folder and the .gitignore?” **Sheldon:** “That is an **infra card**, no triple, just done or not-done.”

...

Sheldon (commit, 2 min): writes the plan into README.md. After the table breaks, **Penny** and **Howard** each **commit a short written review** capturing what they raised. Done.

3.13 Where Planning Lives: Board and Repo

The board is “what is happening now.” The repo is “what happened, in writing.”

```
1 README.md           # Iteration Review per iteration
2 BACKLOG.md          # human-readable mirror of the board
3 data/raw/            # gitignored
4 data/processed/      # committed once it is clean
5 deliverables/        # M1..M5 artifacts (M2 summary lands here)
6 studio/              # briefs + critiques
```

Tonight you commit the **plan** in README.md. After you execute, it fills out into a full Iteration Review (completed PBIs, observations, decisions) before the review studio.

3.14 Plan It: Write One M2 Card

Take three minutes. For your project, write **one** card aimed at the Data Summary, with all the Ready fields.

```
1 Hypothesis: <one sentence>
2 Create: <the artifact you will build>
3 Observe: <the measurement you will record>
4 Analyze: <the decision the measurement lets you make>
5 Tag: M2-data-summary    Size: S / M / L / XL
```

...

Worked example:

- 1 Hypothesis: The transit agency's API has enough coverage to skip the manual export.
- 2 Create: A Python script that pulls all 2020-2024 records into data/raw/ridership/.
- 3 Observe: Count records per year; compare totals to the published annual figures.
- 4 Analyze: If coverage is within 5%, drop the manual export card. If not, keep it.
- 5 Tag: M2-data-summary Size: M

4 Part 3: The Data Story

The Data Summary is not the end of a chore. It is the foundation of the only thing your audience will remember: the story.

4.1 The Data Story

4.2 Start It Now, Not at the Poster

Most teams wait until the poster to think about narrative. By then the data is already collected and the story has to be reverse-engineered out of whatever they happened to gather.

...

Do it backwards on purpose: let the **story you want to tell** shape **which data you prove you have** at M2. The Data Summary and the data story are the same decision viewed from two ends.

4.3 A Data Story Is a User Story

You know this shape from agile. A **user story** is told from the point of view of the person who needs something:

- 1 As a [role], I want [capability], so that [benefit].

...

A **data story** is the same move for analytics, told from the stakeholder who needs the answer, not the analyst who runs the model:

- 1 As a [stakeholder], I need to know [question], so that I can [decision].

That is why DS3 lets every PBI be a hypothesis **or a user story**. Your project is the biggest user story you have; your backlog items ladder up to it.

4.4 The Story Spine

Fill in five blanks and you have a spine.

Beat	Prompt
Context	“For [audience], in a world where [situation]...”
Tension	“...there is a problem: [the gap or pain].”
Question	“So we asked: [your research question].”
Finding	“Using [your data], we found [the insight].”
Stakes	“Which means [the implication / what should change].”

...

You cannot fill the **Finding** beat honestly until your data is in. That is the link back to your M2 gate.

4.5 The Write-Up Skeleton

The spine is the soul; the skeleton is the body it lives in. Every poster and write-up addresses these components. Stub this file in `deliverables/` and grow it.

```
1 # <Project Title>
2 Team, affiliations
3
4 ## Abstract
5 ## Introduction: problem and research question(s)
6 ## Impact and stakeholders
7 ## Related work / background
8 ## Data and data engineering
9 ## Ethics and responsible use
10 ## Methods and evaluation
11 ## Results and visualizations
12 ## Discussion, limitations, and future work
13 ## Conclusions and recommendations
14 ## Reproducibility and references
```

Your project may rename, merge, or add sections, but every write-up addresses these. Start from a `.qmd` so one source renders to HTML and PDF.

4.6 Which Milestone Fills Which Section

Each milestone first-drafts specific sections of that skeleton.

Write-up section	First drafted at
Introduction, problem, research question	M1 Proposal
Impact and stakeholders	M1 Proposal
Data and data engineering	M2 Data Summary
Ethics and responsible use	M1, deepened at M2
Methods and evaluation	M3 Poster Draft
Results and visualizations	M3 Poster Draft
Discussion, limitations, future work	M4 Write-up Draft
Conclusions and recommendations	M4 Write-up Draft
Reproducibility and references	M2, maintained throughout

...

You already wrote two of these at M1. This week you write the next one. M4 is assembly, not authorship from scratch.

5 Wrap

5.1 Tonight and the Next Two Weeks

5.2 What I Want From You

Tonight, in your 20-minute planning studio:

1. **Plan the iteration:** Ready, M2-tagged cards on the board, sized to fit the two-week iteration.
2. **Commit the plan** in `README.md` before the table breaks.

...

Over the next two weeks, toward the Data Summary:

3. **Incorporate proposal feedback** from me and your peer POs into your plan and the Data Summary.
4. **Acquire and organize:** real data landing in `data/processed/`.
5. **Stub the skeleton** and outline your data story.
6. **Submit the Data Summary** by end of week 7, then post your Iteration Review in `README.md` before the week 8 studio.

5.3 References

1. Saltz, J. S., Hotz, N., and Sutherland, A. (2022). *Achieving Lean Data Science Agility Via Data Driven Scrum*. HICSS 55. <https://hdl.handle.net/10125/80218>
2. Data Science Process Alliance. *Data Driven Scrum*. <https://www.datascience-pm.com/data-driven-scrum/>
3. Schwaber, K., and Sutherland, J. *The Scrum Guide*. <https://scrumguides.org/>
4. Anderson, D. J., and Carmichael, A. *The Kanban Guide*. <https://kanbanguide.org/>
5. Knafllic, C. N. (2015). *Storytelling with Data*. Wiley.
6. Duarte, N. (2019). *DataStory: Explain Data and Inspire Action Through Story*. Ideapress Publishing.
7. DATA 510 Data Summary (M2) milestone brief. <https://courses.lpcordova.phd/data510/project/data-summary.html>
8. DATA 510 Project requirements at a glance. <https://courses.lpcordova.phd/data510/project/project-requirements.html>
9. DATA 510 [Studio Session \(DS3 ritual\)](#).