

# **AWS State, Local, and Education Learning Days**

Madison, WI





# How to Become a Data-Driven Public Sector Organization

**Annalisa Mastri, PhD (she/her)**

Data Transformation Strategist  
AWS Worldwide Public Sector  
[aimg@amazon.com](mailto:aimg@amazon.com)

# Agenda

Data-driven public sector organizations

Data & AI strategy components

- Intelligence delivery
- Governance
- Architecture
- Management

Best practices

# Common data challenges



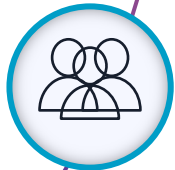
**Business challenges:** Struggle to answer important Qs timely, limited transparency and auditability, limited analytics capacity and capabilities to data



**Data challenges:** No defined governance models, poor data quality, no metadata for audit/compliance/provenance, no data cataloguing



**Operational challenges:** Manual/ad hoc processes consume resources, inconsistent data manipulation, variation in data extraction and identity resolution,



**Financial challenges:** Struggle to address existing technology investments & debt



**Functional challenges:** Not built to scale to new data sources; won't support AI

# Unstructured data is valuable & untapped

## 20% Structured Data

Highly organized and readily accessible, formatted in tables, spreadsheets, or databases with predefined categories and relationships.



Institutional Knowledge

## 80% Unstructured Data

Lack rigid schema and organization, which makes them hard to extract, search, process, and analyze.

**18% of organizations use unstructured data at scale**

# Data is critical to AI/ML applications

“The opportunity is vast, and we need to move fast.  
We need to scale strategically, data matters.”

93%

realize a data strategy is  
crucial to getting  
business value from  
generative AI

37%

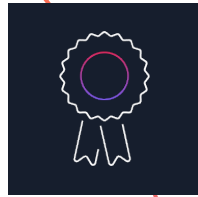
have the right data  
foundation to pivot to  
generative AI

46%

see data quality as the  
biggest blocker

Source: AWS, 2024 CDO Insights: Data & Generative AI  
Slide credit: Mary Strain

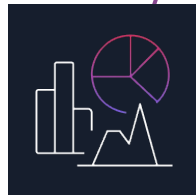
# What we can achieve when we are data-driven



**Insights at your fingertips** through real-time monitoring, automated reporting, secure data sharing



**More productive program and IT staff** who do less manual data entry, data transformation, compliance activities, reporting

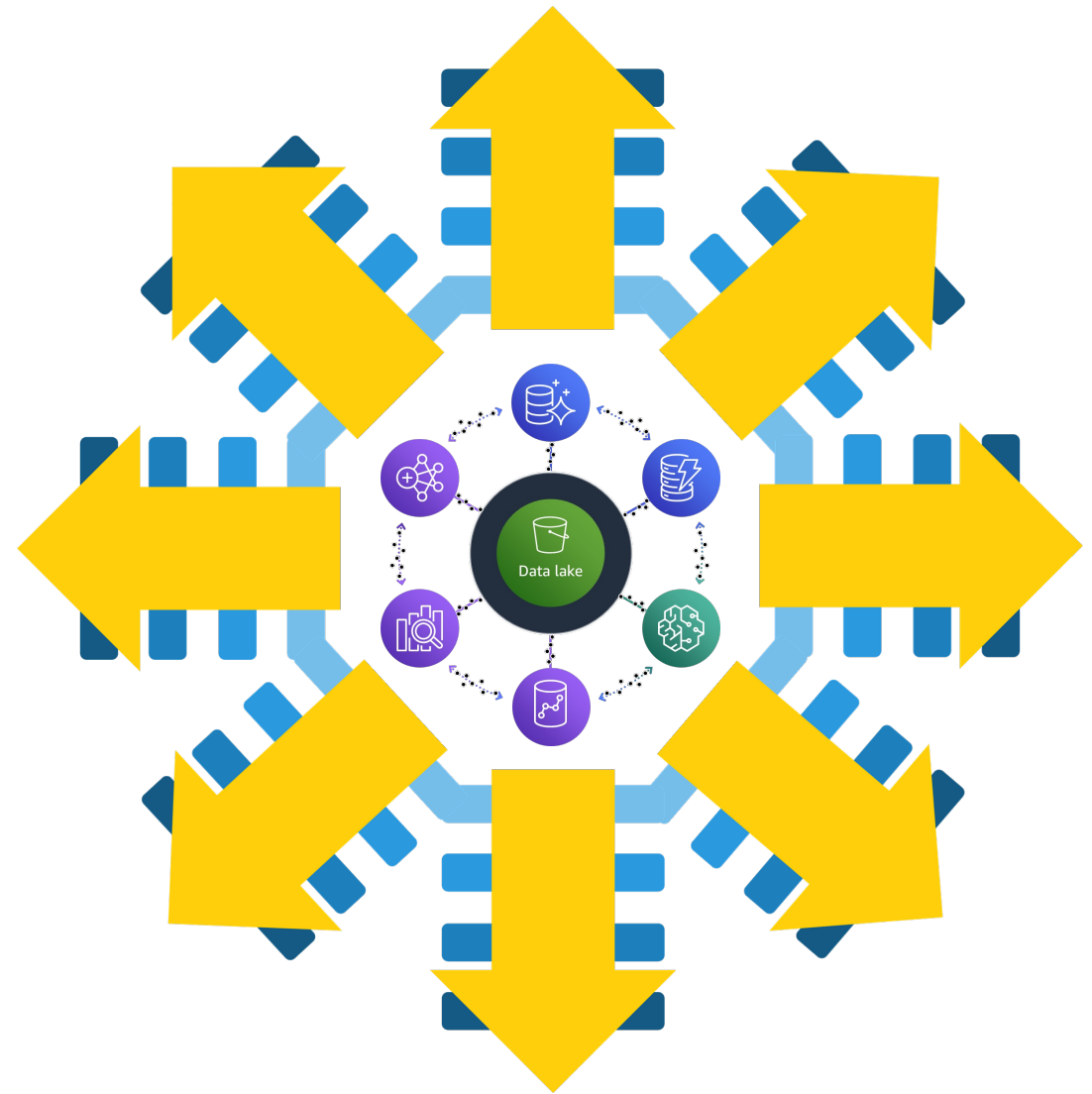


**Empowered teams** that focus on outcomes, make data-driven decisions, meet legal and policy requirements

# What is a data-driven public sector organization?

- Data is valued as a **strategic asset**, not just a byproduct of business processes
- Data is **FAIR** (Findable, Accessible, Interoperable, and Reusable)
- Stakeholders throughout the organization are **empowered** to make decisions

**Supporting  
multiple use cases  
exponentially  
increases the value  
of your data**



# How do we become data-driven?

---

## Culture

Embed data-driven decision-making seamlessly into operations

---

## People and process

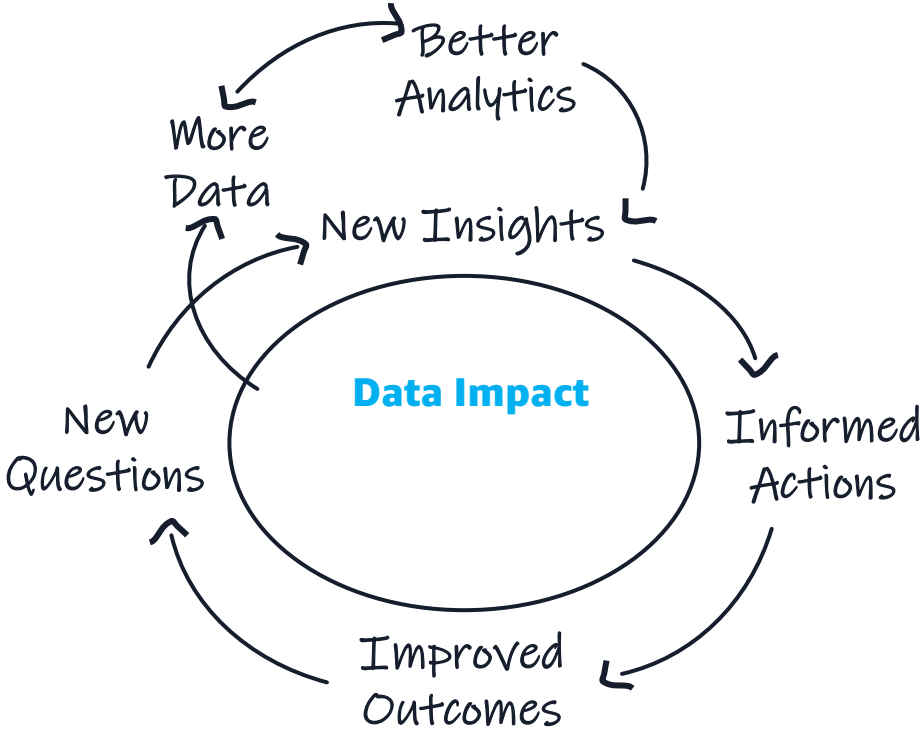
Build the right process models and ensure staff have needed skills

---

## Technology

Select technology solutions that serve defined business use cases

# The Data-Driven Organization



“An organization that harnesses data as an asset to drive sustained innovation and create actionable insights that improve decision-making about outcomes constituents care about, thereby building greater trust.”

## Key Drivers



Mindset and culture



Data strategy



Business-IT alignment



Strong collaboration across data producers and consumers



Empowered producers and consumers

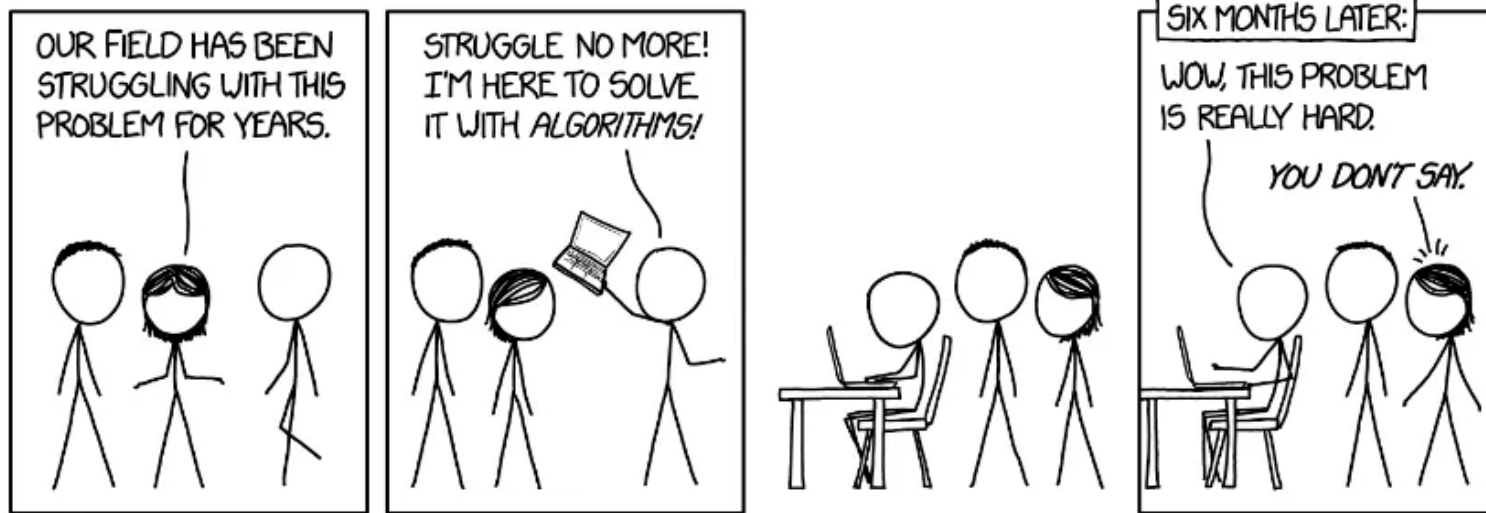


Nimble procurement processes





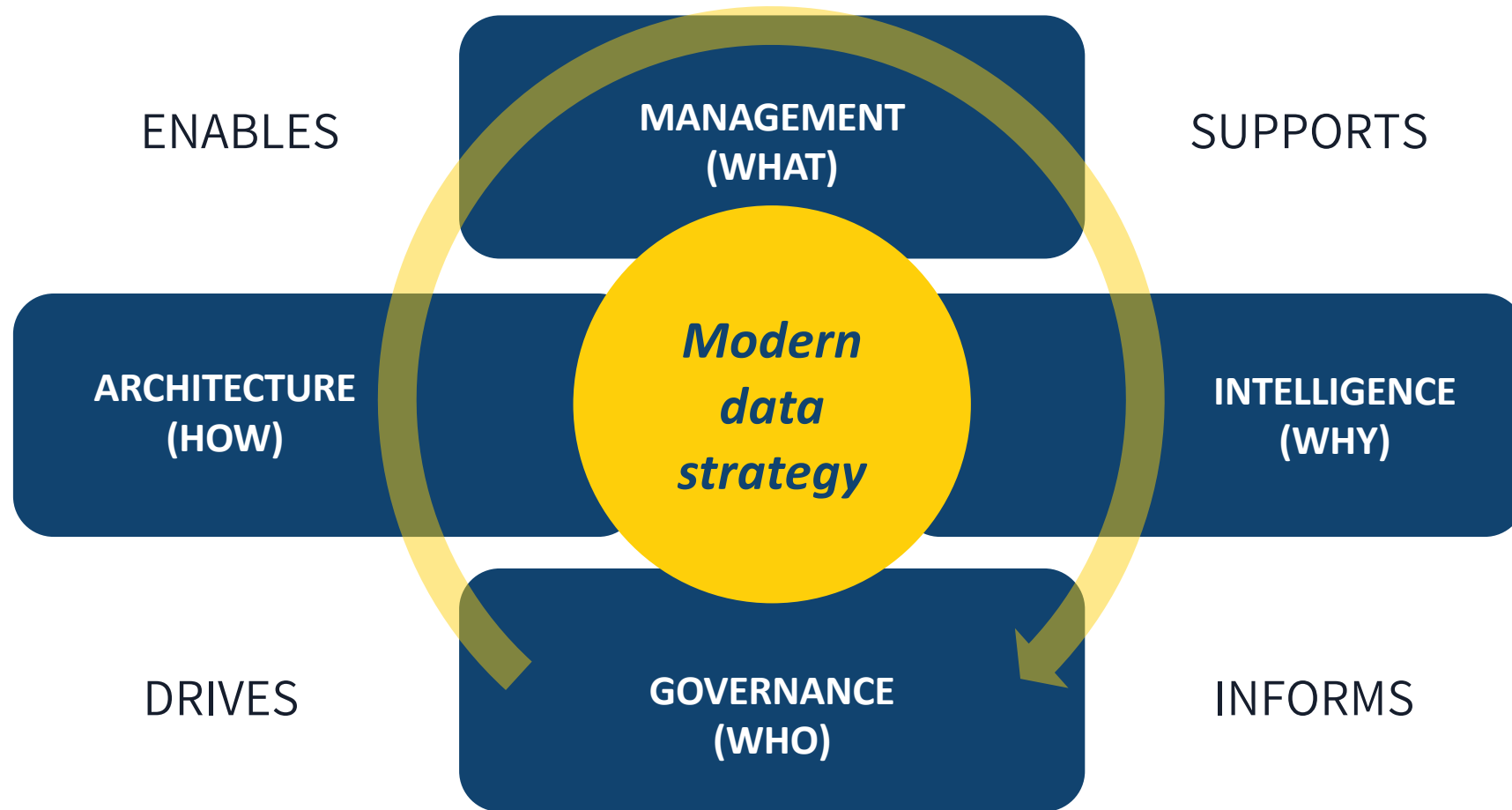
**we are not building one of these...**

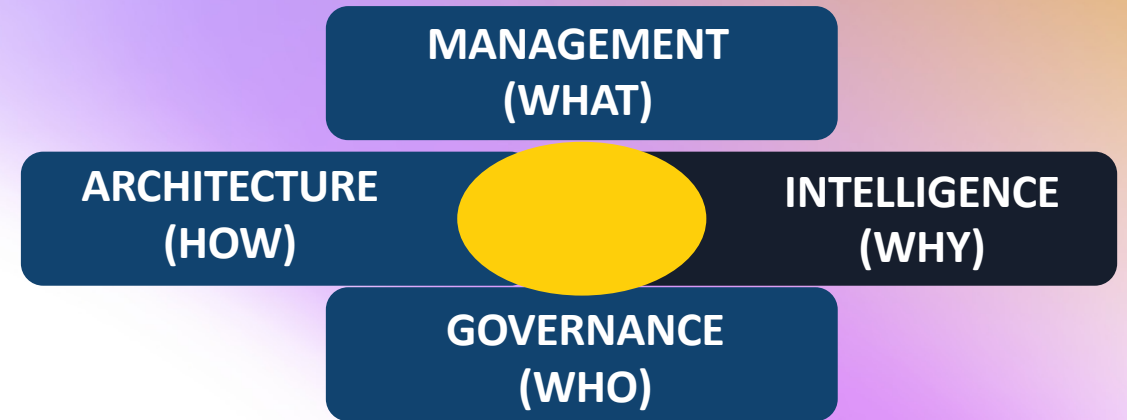


Credit: XKCD.com

# Modern data and AI strategy

# A modern data strategy is more than tech



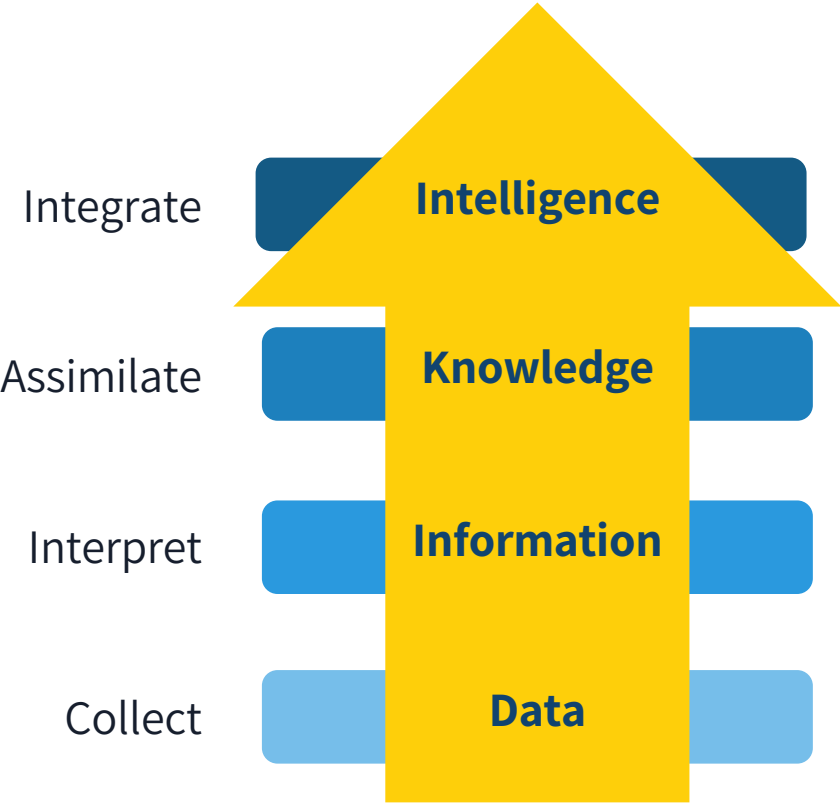


# Modern data intelligence

## Why we collect and manage data

# Transforming data into intelligence increases its value

## Actionable decisions



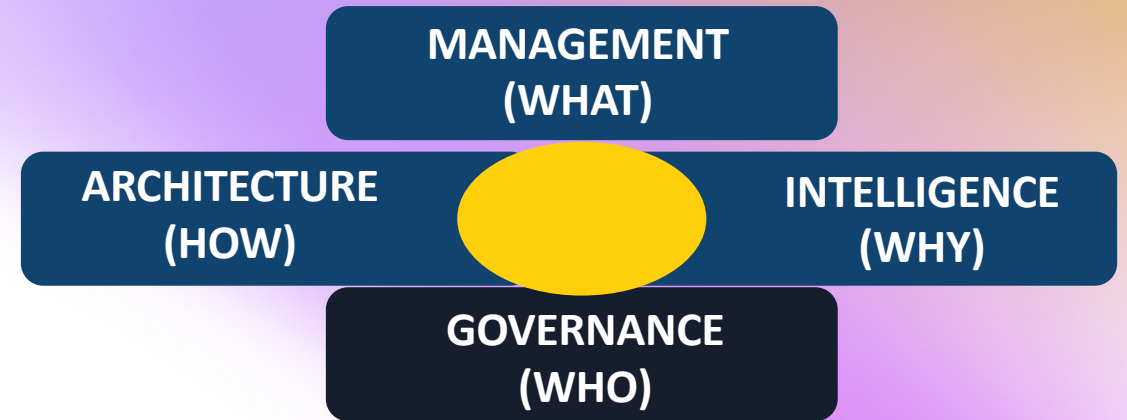
# Start with 'Why'

Improve **operational efficiency**

Provide more **customer-centered services**

Achieve **better outcomes** for constituents

Answer **important questions** from stakeholders



# Modern data governance

Who can do what with data

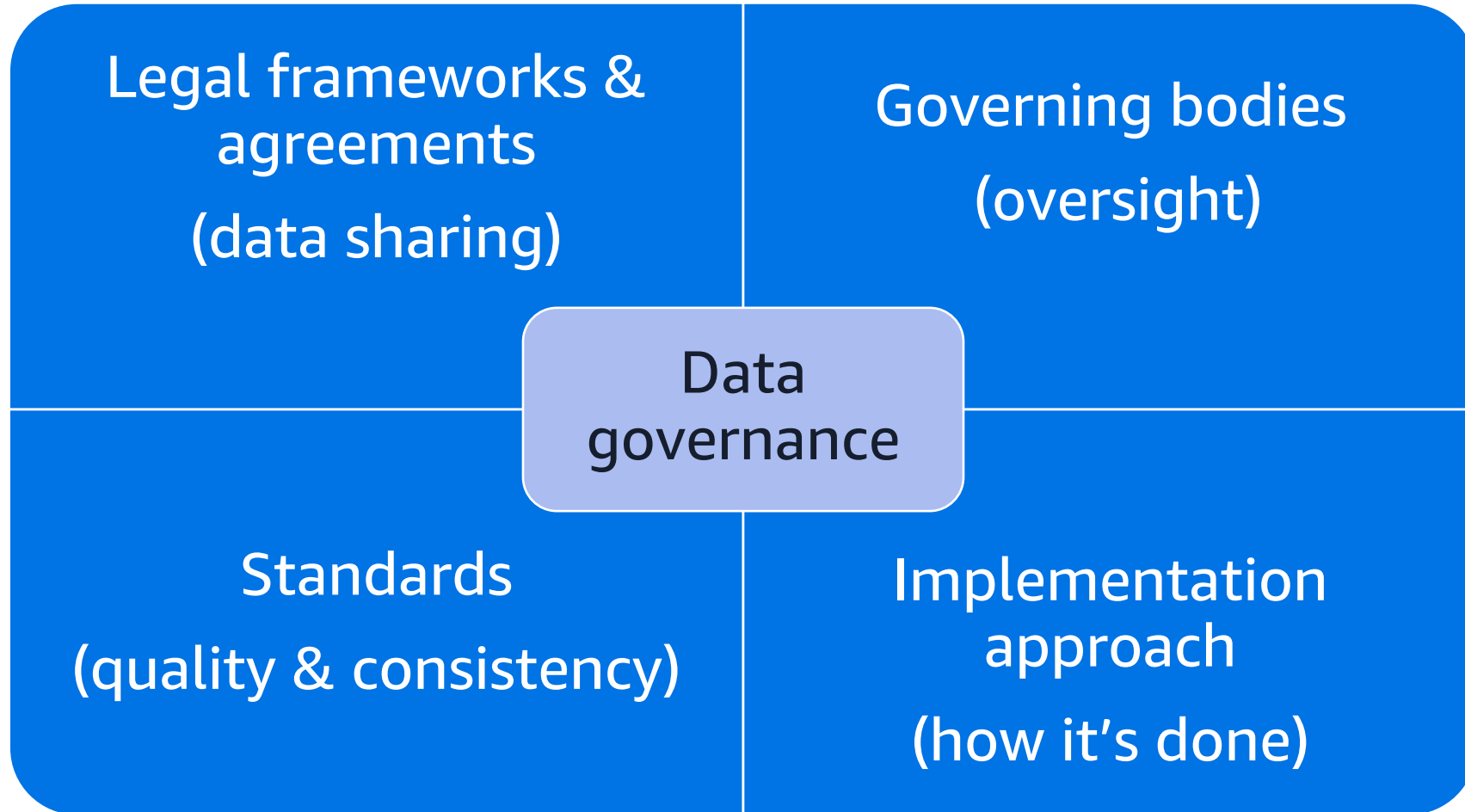
# What is data governance?

Data governance is the combination of **people, processes, and technology** that organizations use to ensure the quality and security of their data throughout its lifecycle

Creates accountability through clearly defined roles and responsibilities

Builds trust among stakeholders by balancing the desire for data access with the necessity of privacy and security

# Data governance building blocks



# Data sharing models

How accountability for data sharing happens



# Ad hoc approach: Direct agency-to-agency sharing, customized DSAs

## Advantages

- Maximum flexibility
- Can start immediately
- Clear accountability between parties

## Disadvantages

- Time-consuming & resource-intensive
- Inconsistency across agreements
- Difficult to track and ensure compliance

# Templated data sharing: standardized DSAs for agency-to-agency sharing

## Advantages

- More efficient than ad hoc
- Consistent terms and controls
- Predictable process for agencies

## Disadvantages

- Tracking individual agreements
- Templates may not fit all scenarios
- Risk of “set it and forget it”

# Federated data sharing: central coordination body with distributed control

## Advantages

- Agencies retain control of data and sharing decisions, DSAs
- Shared standards & policies → consistency
- Coordination body facilitates but doesn't control

## Disadvantages

- Requires sustained commitment
- Can be slow to implement changes across all agencies

# Data Trust: centralized entity manages all sharing; data contribution agreement between producer and trust, data use agreement between consumer and trust

## Advantages

- Single point of control → consistent policy enforcement
- Clear accountability for decisions
- Streamlined processes for managing access & compliance

## Disadvantages

- Can become a bottleneck
- Requires dedicated operational support and resources
- Agencies may resist loss of direct control

# Comparing data sharing models

Model	Control	Speed to implement	Resource needs	Risk level	Scalability
Ad hoc	Agency	Fast	Low	High	Low
Templated	Agency	Medium	Low-med	Medium-high	Medium
Federated	Agency + standards	Slow	Medium	Medium	High
Data trust	Central trustee	Slow	High	Low	High

# Governance roles and structures

How data sharing is overseen



# Key governance roles

**Data owners** are senior officials with accountability and decision-making authority over specific data assets within their domain.

**Data stewards** are subject matter experts who manage data quality and implement governance policies in day-to-day operations.

**Data custodians** are technical staff responsible for the secure storage, maintenance, and accessibility of data systems.

# Data governance structures oversee data operations



## Advisory Board

- High-level leaders (people business leaders report to)**
- Set strategic direction and vision for DG
  - Prioritize DG initiatives for implementation
  - Provide executive sponsorship for major data initiatives

## Executive Steering Committee

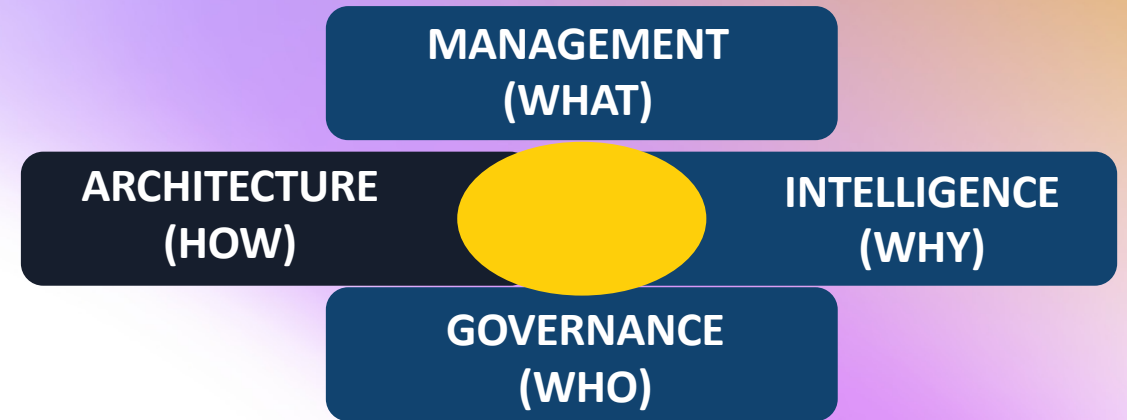
- Dept./ agency heads (people data owners report to)**
- Ensure DG aligns with business objectives
  - Allocate resources to support DG initiatives
  - Resolve escalated issues

## Data Governance Council

- Data owners, CDO, Privacy & Security Officers**
- Approve data governance policies, standards, and best practices
  - Address cross-cutting data governance issues
  - Oversee data quality, security, compliance



**Establishing governance bodies is step one. Steps two through one hundred are where the real work happens.**



# Modern data architecture

How data is stored and processed

# Modern data architecture principles

Cloud-forward

Security and governance

Component architecture

Decoupling

Purpose-built, scalable storage

Transformation flexibility

Data sharing & interoperability

Automation & orchestration



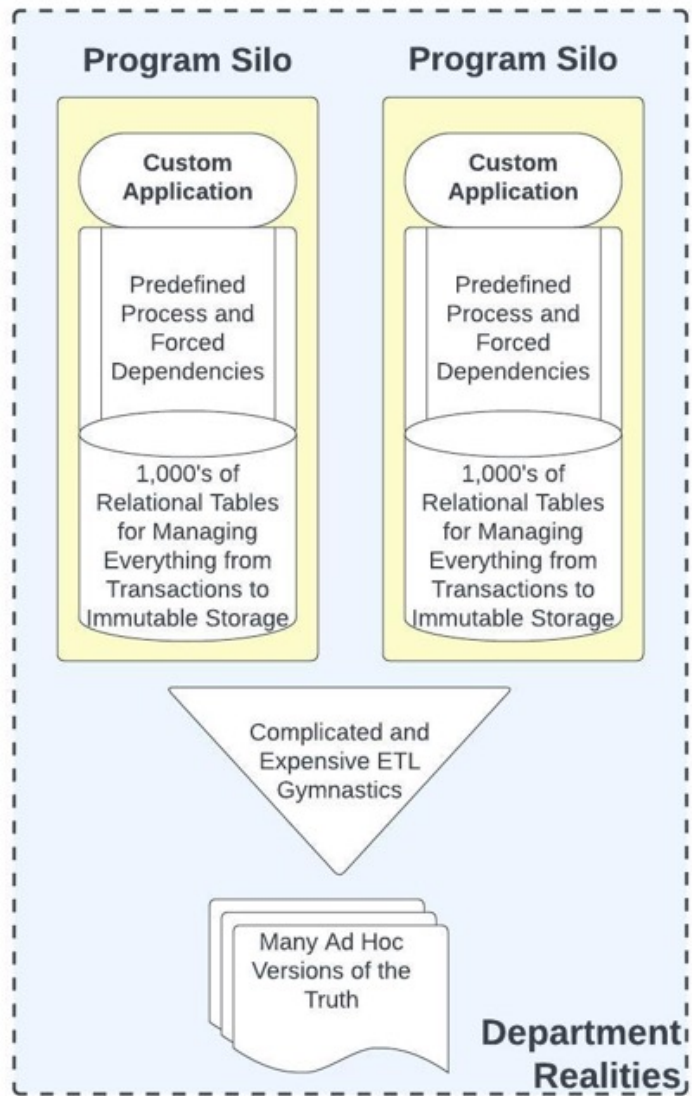
# Anticipation and intentionality

Modern data architecture allows a building block approach with staging and whatever speed you want

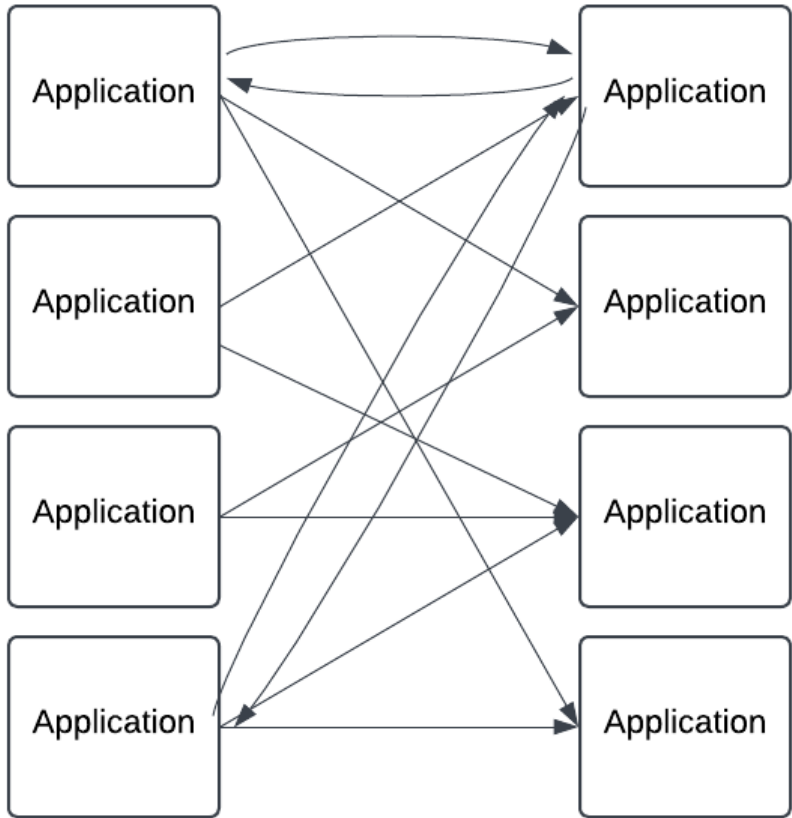
Getting started can be low risk by using best practices

Full transformation takes time, but modernization projects for specific use cases can start right away and scale quickly

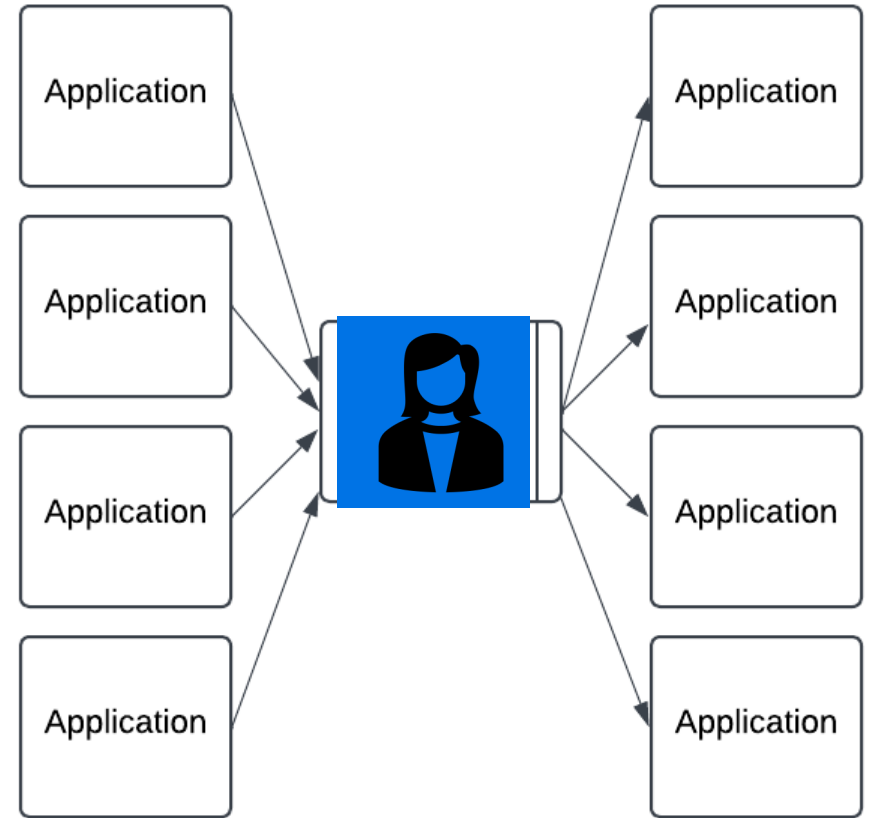
# Today



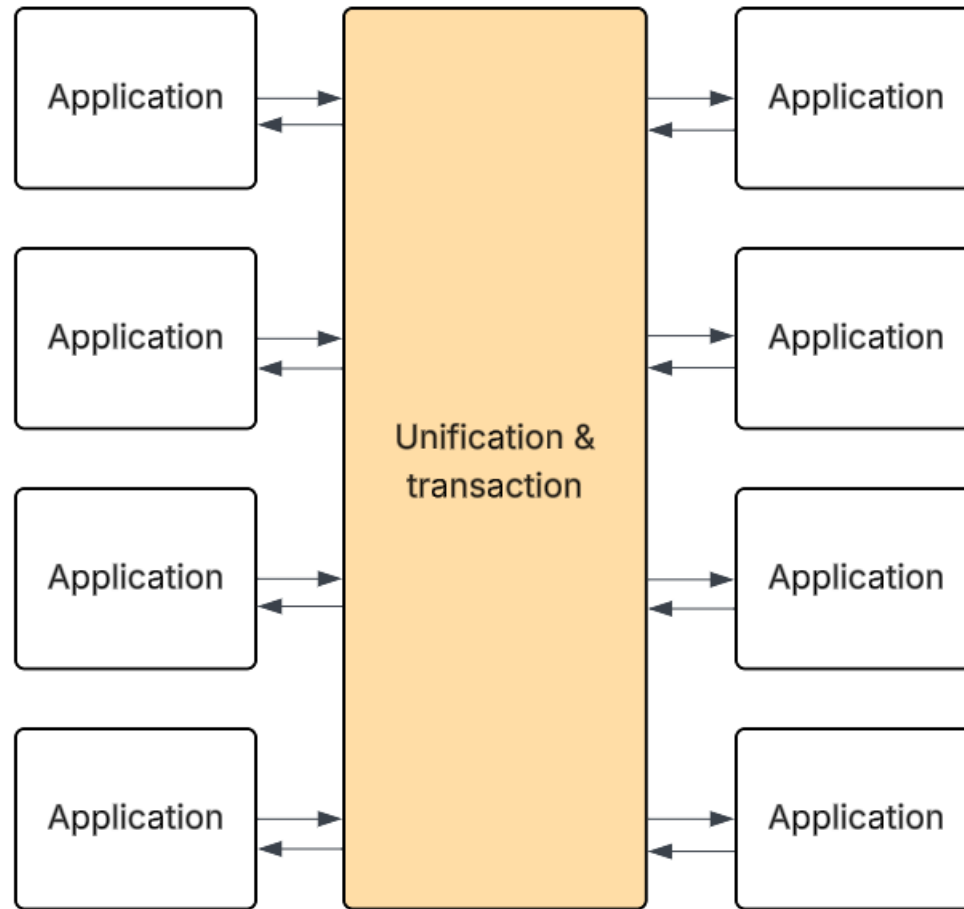
## Point-to-point mapping



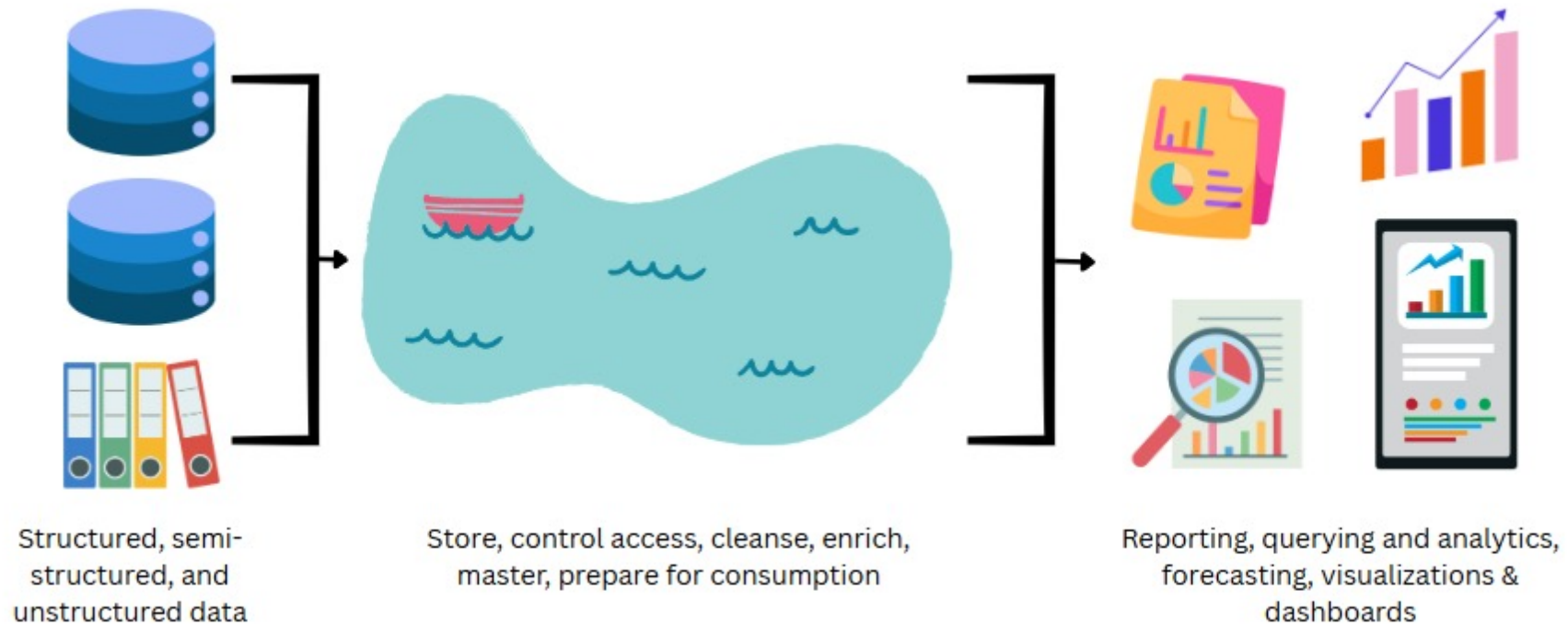
## Manual effort



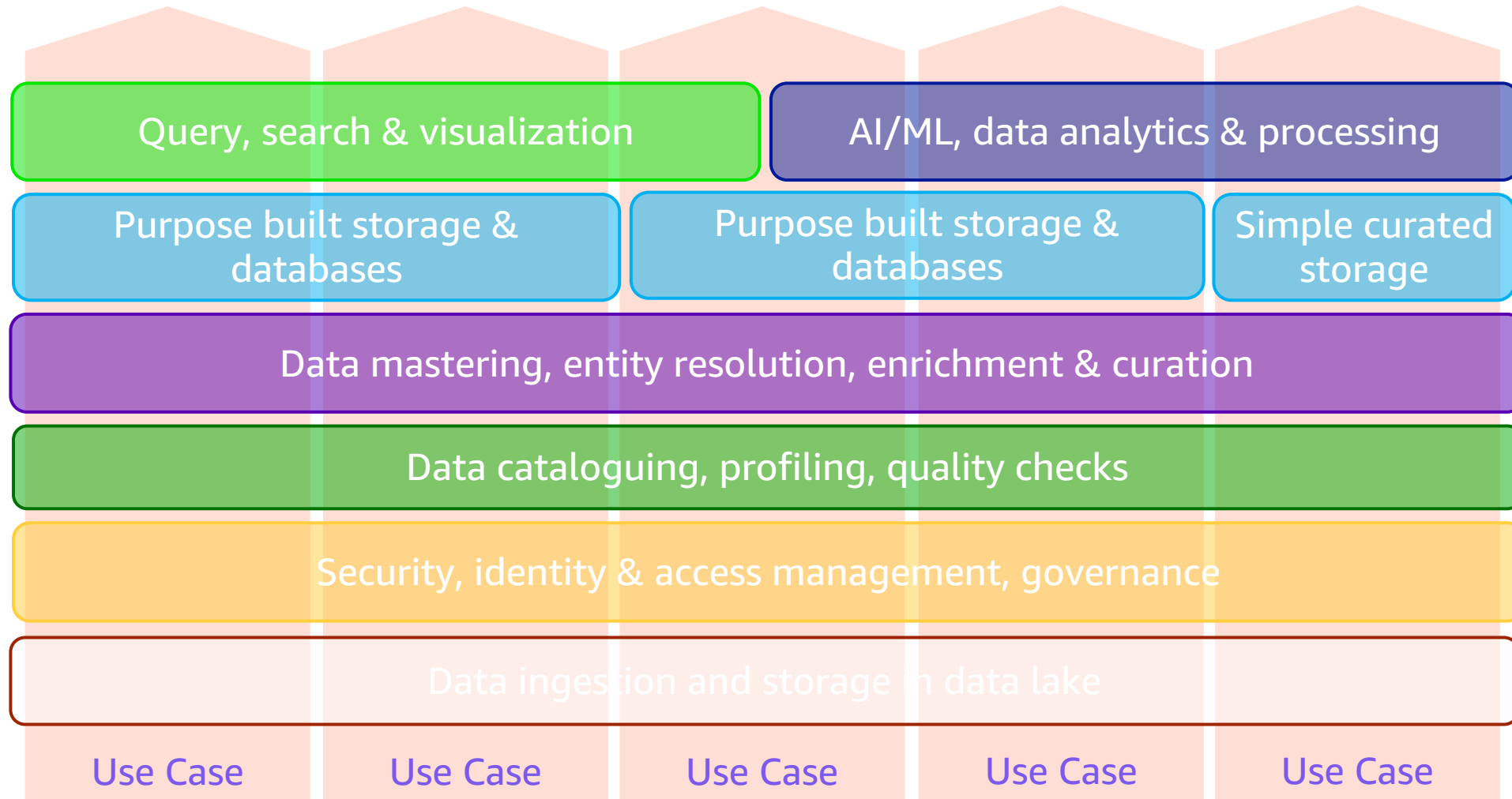
# Modern data architecture inserts a layer between front end and back end

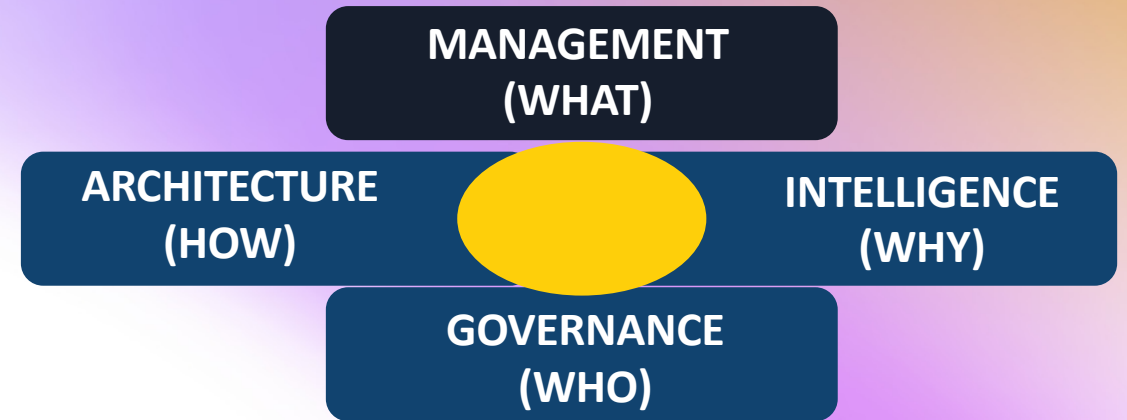


# Data lake: flexible storage



# A modern data architecture is more than a data lake

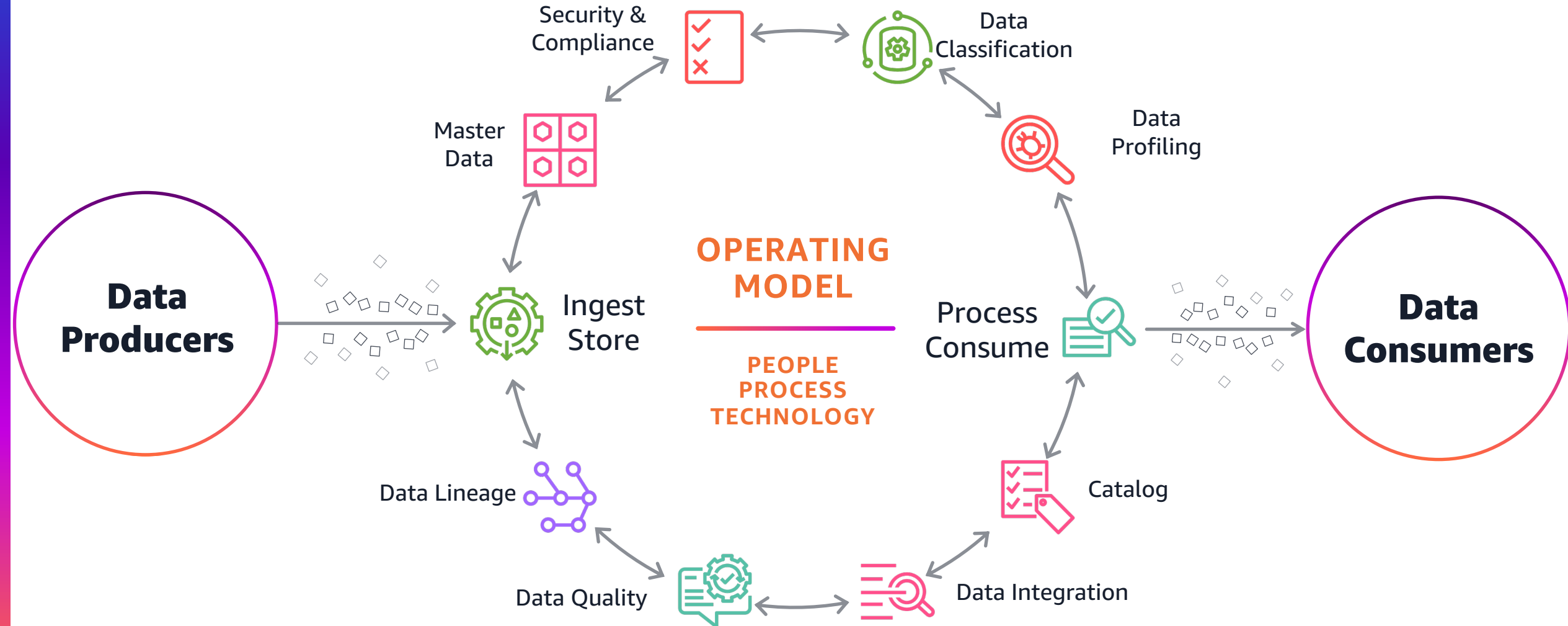




# Modern data management

What happens to data throughout its lifecycle

# Modern data architecture supports data management



# Data and AI strategy best practices



# Mission use cases build the path to data program maturity

## FAACT – Framework for Addiction Analysis and Community Transformation

Data trust, governance framework, data sharing platform, data integration services, and community-based intelligence products

1

## Virginia Workforce Referral Portal

+ Real-time integration between eight production data systems

2

## COVID-19 Pandemic Response

+ Private data trust members, strategic level stakeholders

3

## COVID-19 Vaccine Administration

+ Enterprise level entity resolution service integrating five data systems in near real-time, dissemination of PII to health districts and local health departments for vaccine appointment scheduling

4

## Human Trafficking Analytics

+ AI automating the ingestion of national data from disparate (non-trust member) sources

5

# Do this...

Start with a concrete use case that *drives mission value*

Have an enterprise-wide vision (“think big”) in mind, but **start small**

**Build** data and AI capabilities **over time**

**Align** data and AI strategy **with existing strategic planning process**, so it is continually revisited and revised and in alignment with business initiatives

**Capture success stories** and **share them broadly** to get more stakeholders interested in becoming data-driven

# Don't do this...

Bite off more than you can chew

Select initial use cases with limited mission value

Use AI for the sake of saying you've done it

Create overly bureaucratic processes

Invest in technology tools before defining key policies and workflows

# Data-Driven Transformation resources

# Data-Driven Transformation approach



## Inspire

Dive deep on modern approaches and best practices in the field



## Assess

Map where you are and where you're headed



## Enable

Evaluate options and understand multiple paths forward



## Empower

Choose what's right for you and the best way to sustain delivery

# Data Transformation Accelerators

---

## Data Transformation Foundation

*90 minutes*

Learn about data transformation, and the agencies and institutions who have built modern data environments.

---

## Data & AI Strategy Jumpstart

*4 Hours*

Start building your strategy to transform raw information into actionable insights and accelerate AI adoption.

---

## Data Governance Accelerator

*6-8 hours*

Define a data governance approach and start developing the necessary structures, roles, and processes.

---

## Data Architecture Blueprint

*4 hours + follow-up*

Map your current data environment to a modern future state via a collaborative architectural assessment.

# Questions?



# Additional resources



## Learn More!

Learn how data modernization supports government efficiency.



## Explore!

Explore how Virginia used data to become more resilient.



## Connect with D2T!

Reach out to get started on your data transformation journey.



# Thank you!

**Annalisa Mastri, PhD (she/her)**

Data Transformation Strategist  
AWS Worldwide Public Sector

[aimg@amazon.com](mailto:aimg@amazon.com)

**Please complete the survey  
for this session**



**Executive track**

How to become a data-driven  
public sector organization

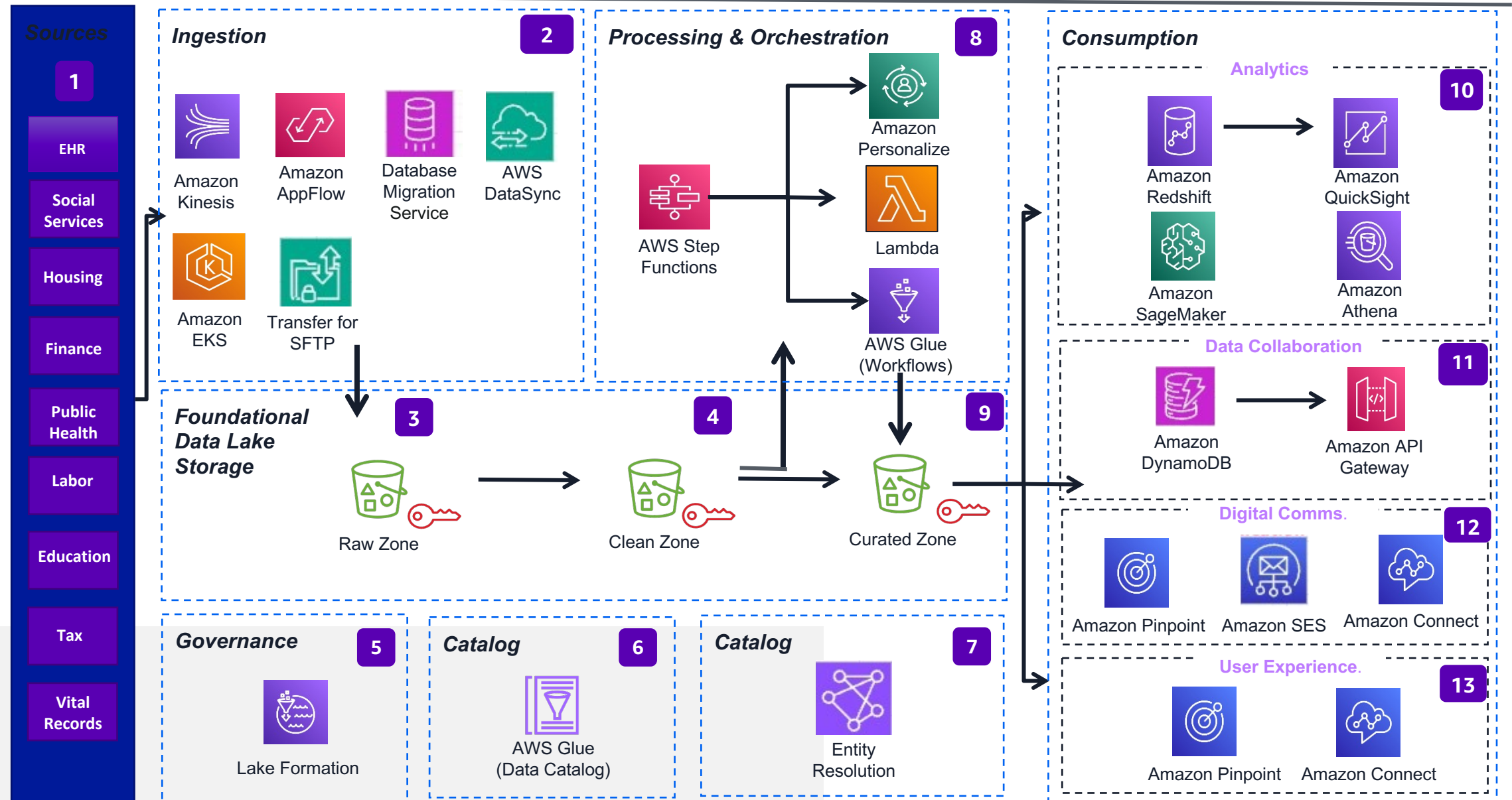
# Extra slides

# Why modernize?

- Repeatable—build it once and use it over and over
  - Saves time and cost
- Resilient, available, and scalable
- Empowered data owners, producers, and consumers
  - Data consumers can get the right information at the right time from the right sources
- **Increase trust in data**
- Take advantage of AI/ML
  - AIML use cases depend on high quality data
  - Increase operational efficiency
  - Take advantage of unstructured and semi-structured data
  - Reduce manual effort
  - Use predictive analytics
- Build for change
  - When the next tech advances come along, they are easy to swap in

# Modern Data Architecture

This guidance provides a reference architecture showing best practices in the building of a customer data platform covering data ingestion, identity resolution, segmentation, analysis and activation.



# Modern data facilitators maturity model

	1	2	3	4
MINDSET AND CULTURE	Data is seen as IT's responsibility. Teams operate in silos. Data governance is informal or non-existent, and there's no sense of a data culture.	Data's broader value beyond IT is emerging. Some enterprise-wide thinking appears, and basic governance structures are introduced. While still largely project-focused, data culture is growing.	Data is recognized as a critical asset, with most teams thinking enterprise-wide. Formal data governance is widely implemented, and a data culture is embraced in most areas.	Data culture is fully embedded in the organization's DNA. Enterprise-wide thinking and governance are the norm. Data-driven decision-making is seamlessly integrated into operations.
LEADERSHIP CHAMPIONS	Buy-in from key IT, data, and business leaders does not yet exist.	Buy-in from key IT, data, and business leaders exists.	Key IT, data, and business leaders own and champion data modernization efforts.	Leaders and staff throughout the enterprise own, champion, and continually reinforce the importance of ongoing data modernization efforts.
SKILLED WORKFORCE	IT and data staff need to upskill considerably to undertake data modernization efforts.	IT and data staff are upskilling for data modernization but it remains difficult to hire and/or promote into key positions.	New IT and data staff roles and career pathways are established; key positions are filled.	IT and data staff roles and career pathways are well documented; key positions are filled. Ongoing professional development opportunities for staff are encouraged and supported.
COLLABORATION	Limited collaboration across business, data, and IT and across departments.	Cross-functional and cross-department teams and owners are identified but few or no roadmaps or frameworks exist for them to follow.	Roadmaps and frameworks have been developed and guide the work of cross-functional and cross-department teams. IT, data, and business leaders and staff communicate regularly.	Processes and routines for ongoing review and revision of roadmaps and frameworks are in place and being followed. IT, data, and business leaders collaborate to make IT- and data-related decisions.
DATA STRATEGY	Data strategy has not yet been developed.	Data strategy exists for a few mission-critical use cases.	Data strategy exists for all mission-critical use cases.	Data strategy exists for all mission-critical use cases, and processes exist to implement, continuously monitor, and update the strategy.
IT PROCUREMENT	Monoliths still procured through traditional procurement processes.	Moving away from monolithic procurement, but procurement processes remain a barrier to adopting utility-based IT infrastructure.	Procurement processes are more flexible and enable adoption of utility-based IT infrastructure.	IT procurement processes are iterative and nimble. Enterprise has criteria and tenets that drive investment decisions.

# Modern data maturity model: built for change

Implementation driver	What it means to be built for change
Mindset & culture	Data culture is fully embedded in the organization's DNA. Enterprise-wide thinking and governance are the norm. All decisions are driven by insights from data.
Leadership champions	Leaders throughout the enterprise own, champion, and continually reinforce the importance of ongoing data modernization efforts.
Skilled workforce	IT and data staff roles and career pathways are well documented. Key positions are filled. Staff receive ongoing professional development to keep skills sharp.
Collaboration	IT, data, and business leaders collaborate to make IT- and data-related decisions. Processes and routines for ongoing collaboration are in place and being followed.
Data strategy	Data strategy exists for mission-critical use cases, and processes exist to implement, continuously monitor, and update the strategy.
IT procurement	IT procurement processes are iterative and nimble. Enterprise has criteria and tenets that drive investment decisions.





















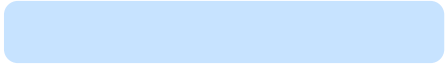




















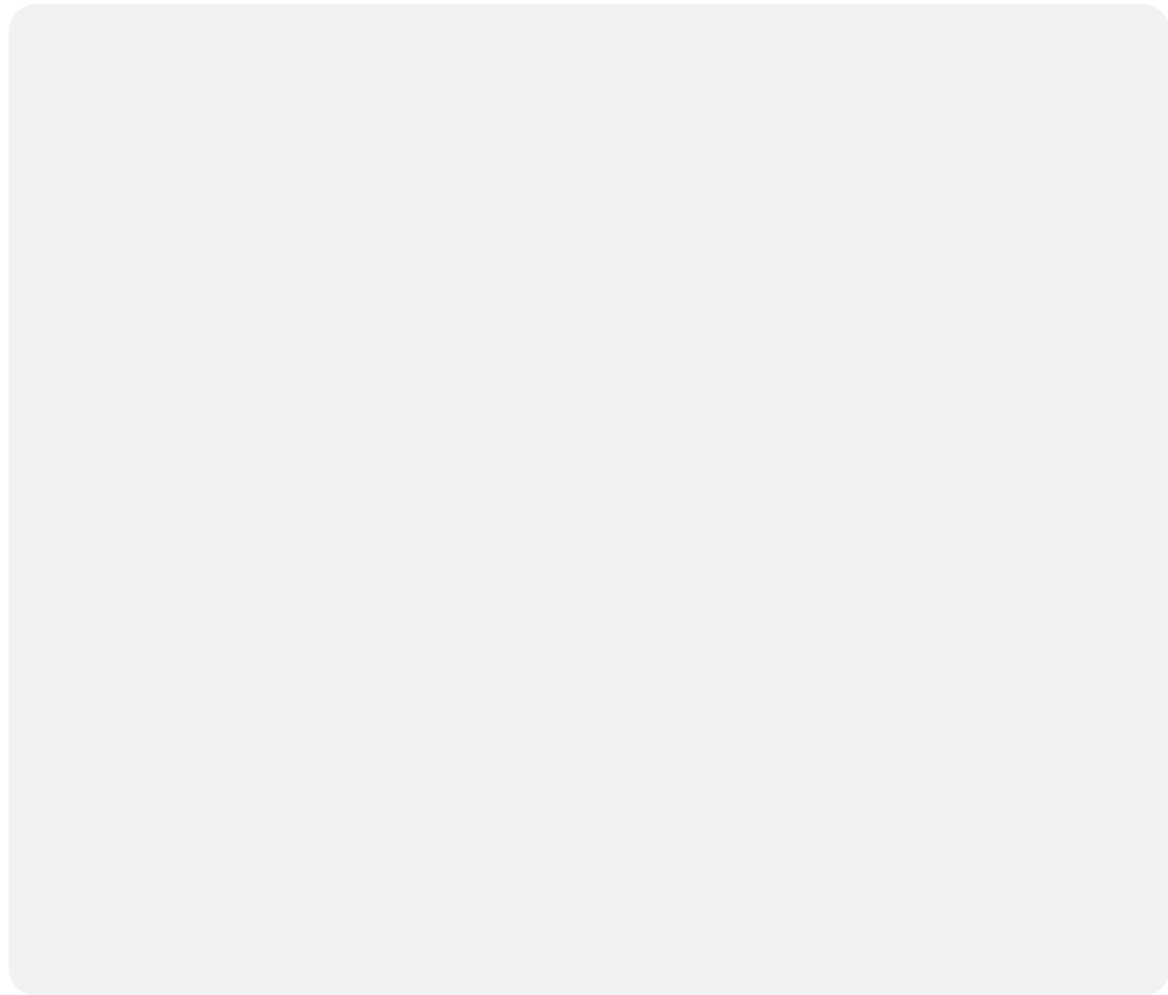
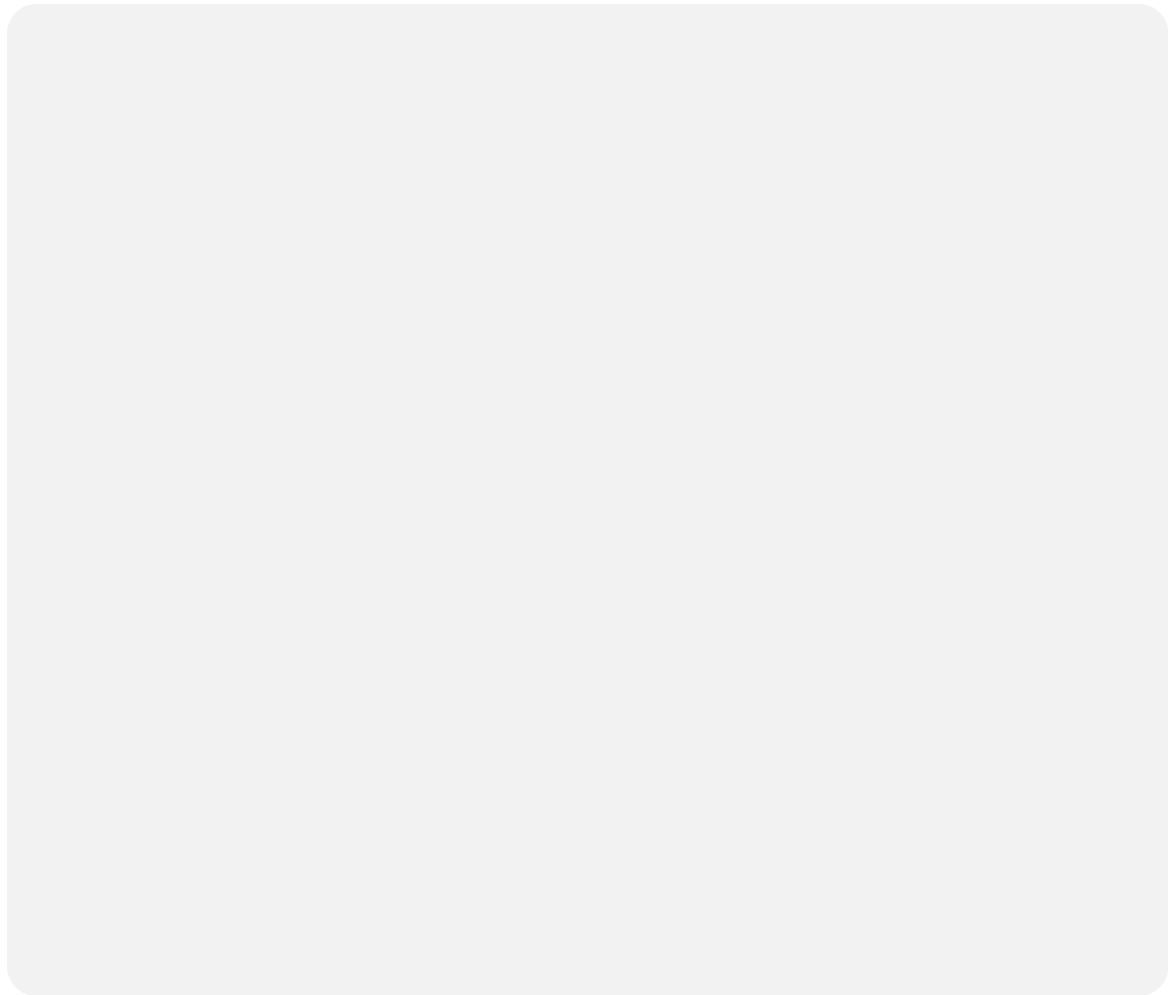






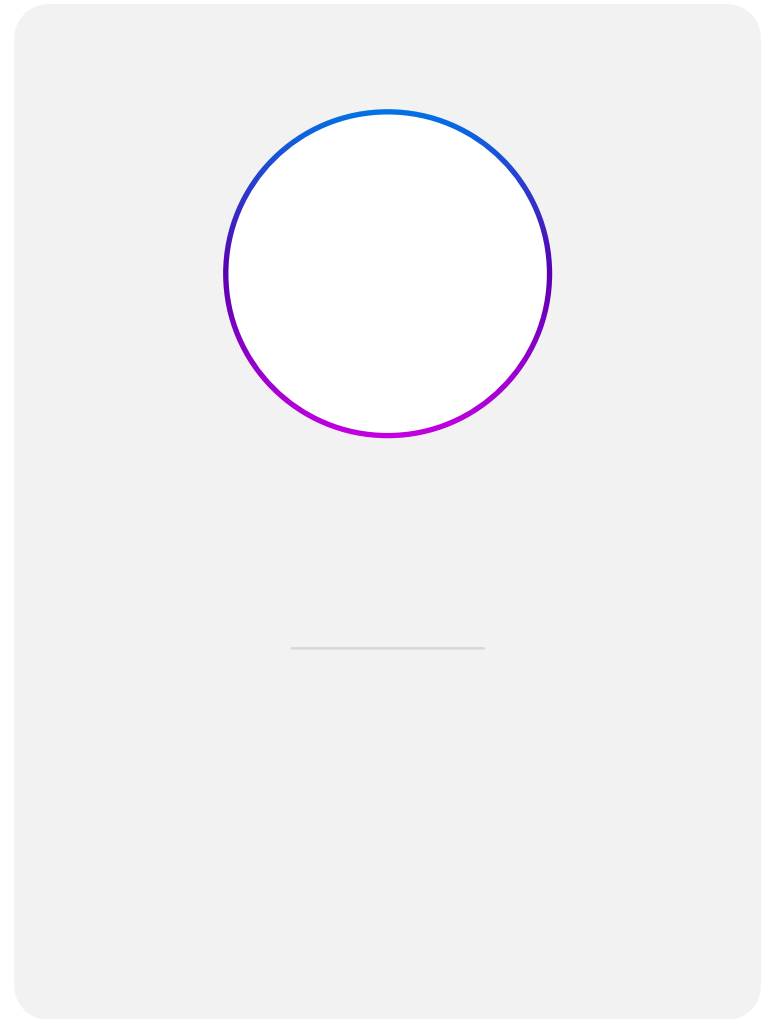
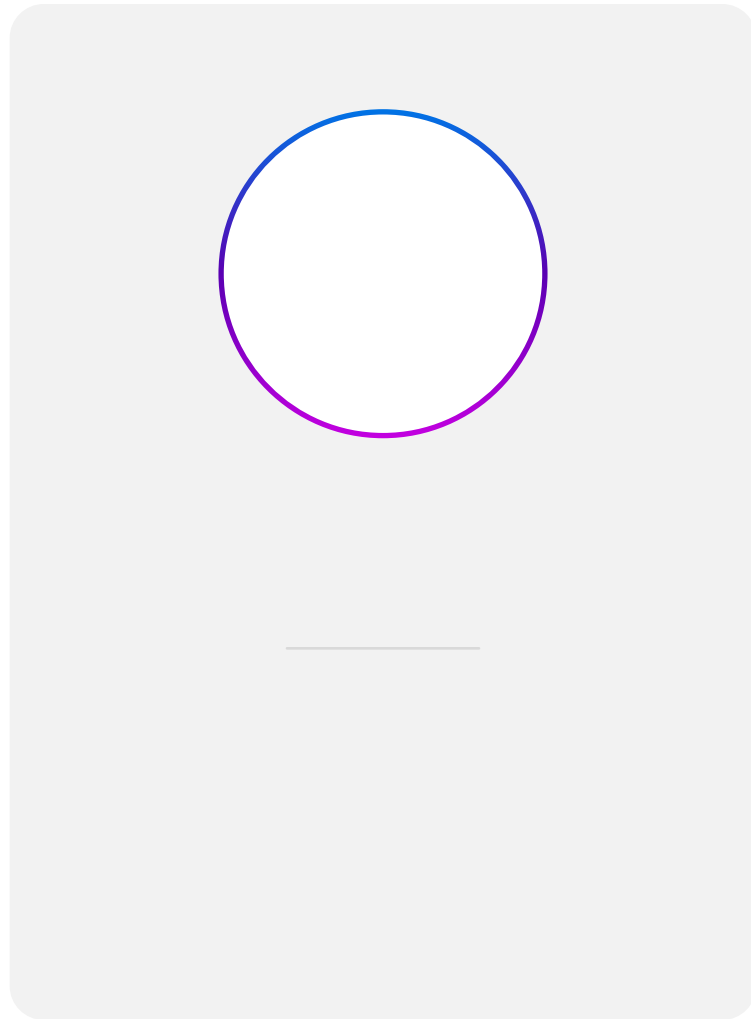
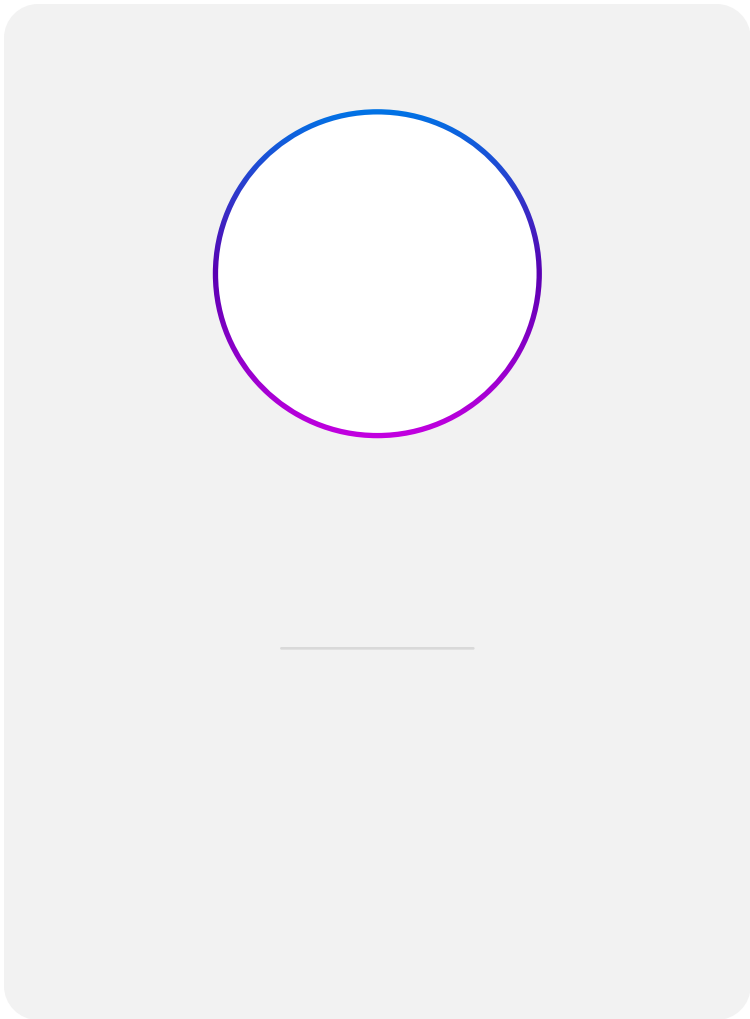


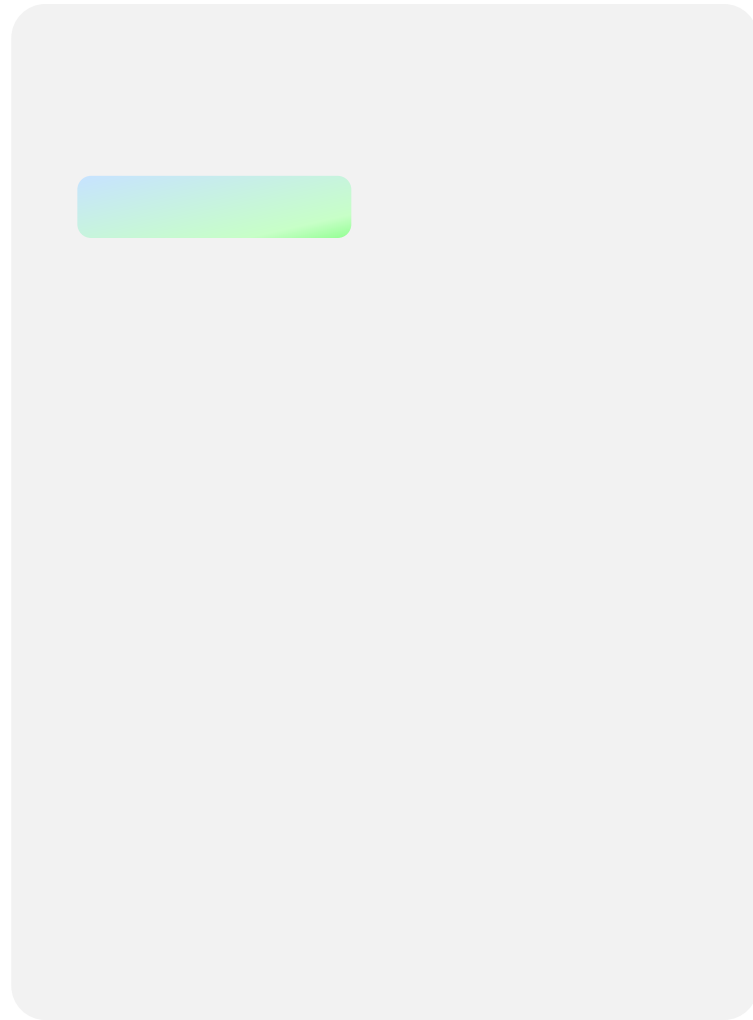
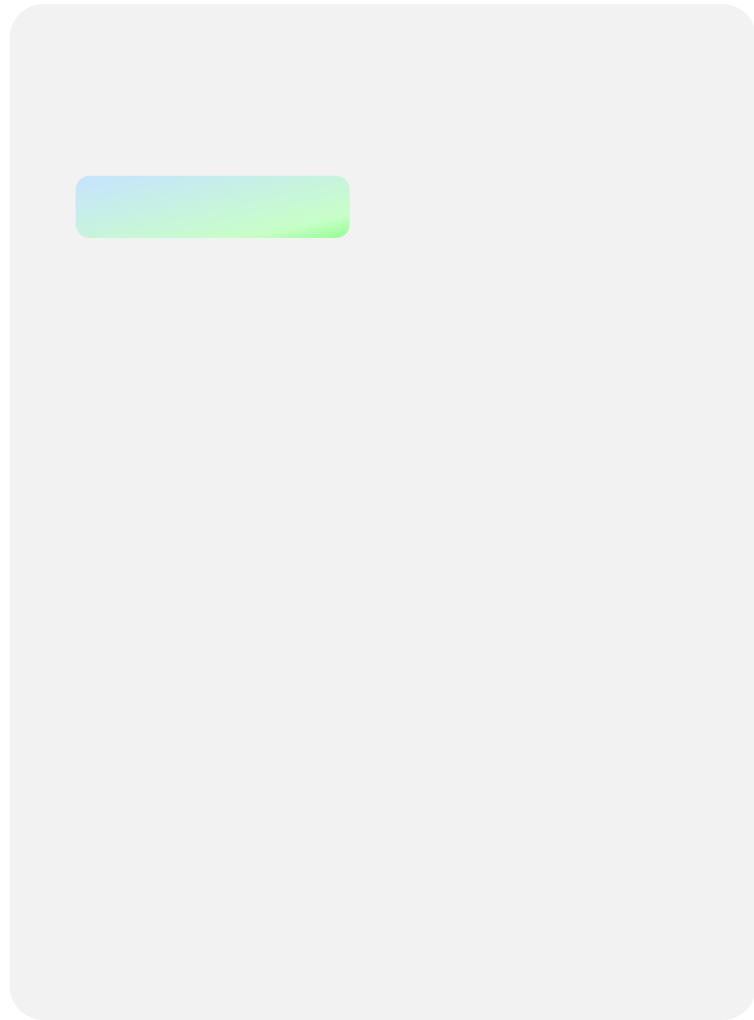
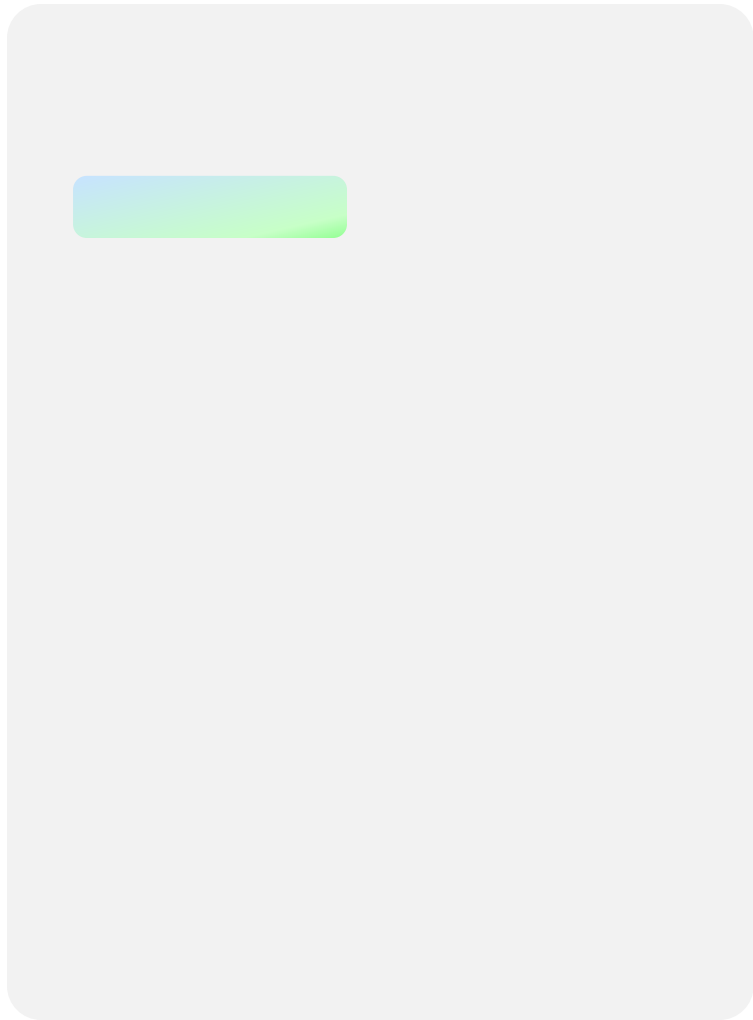


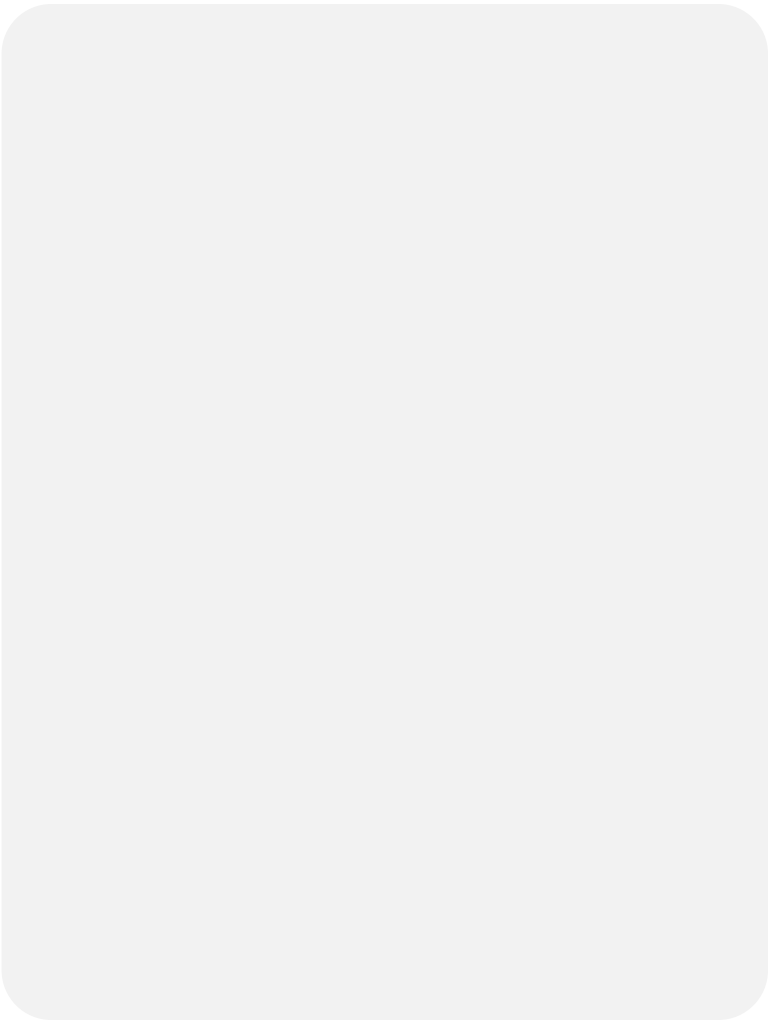
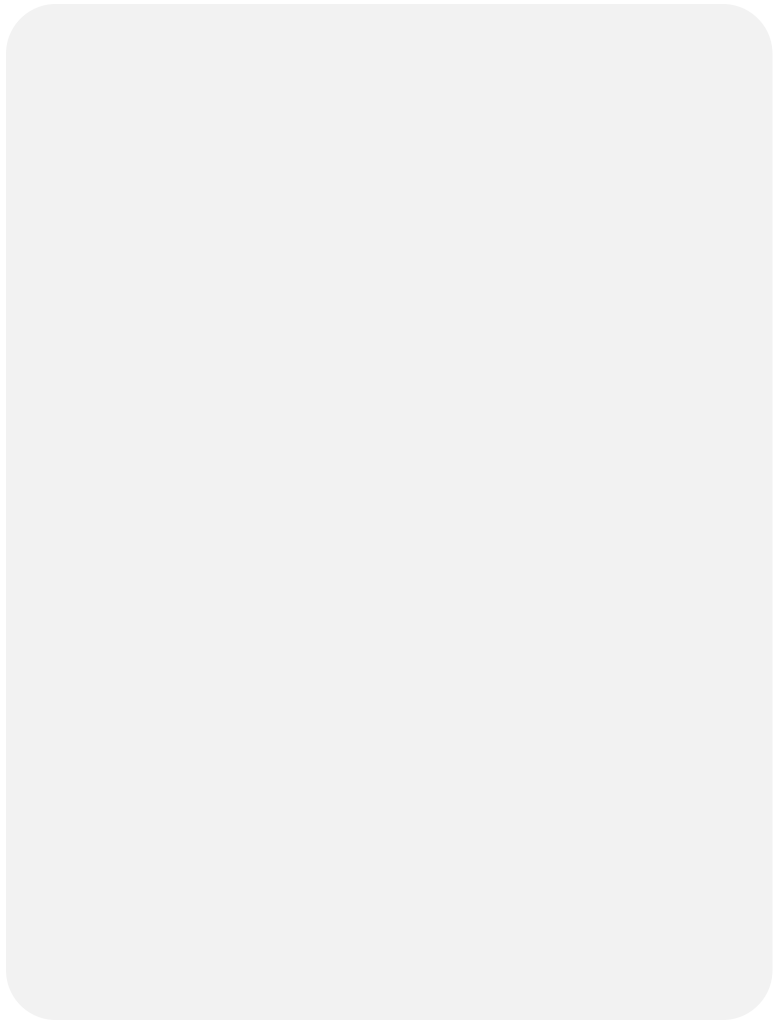
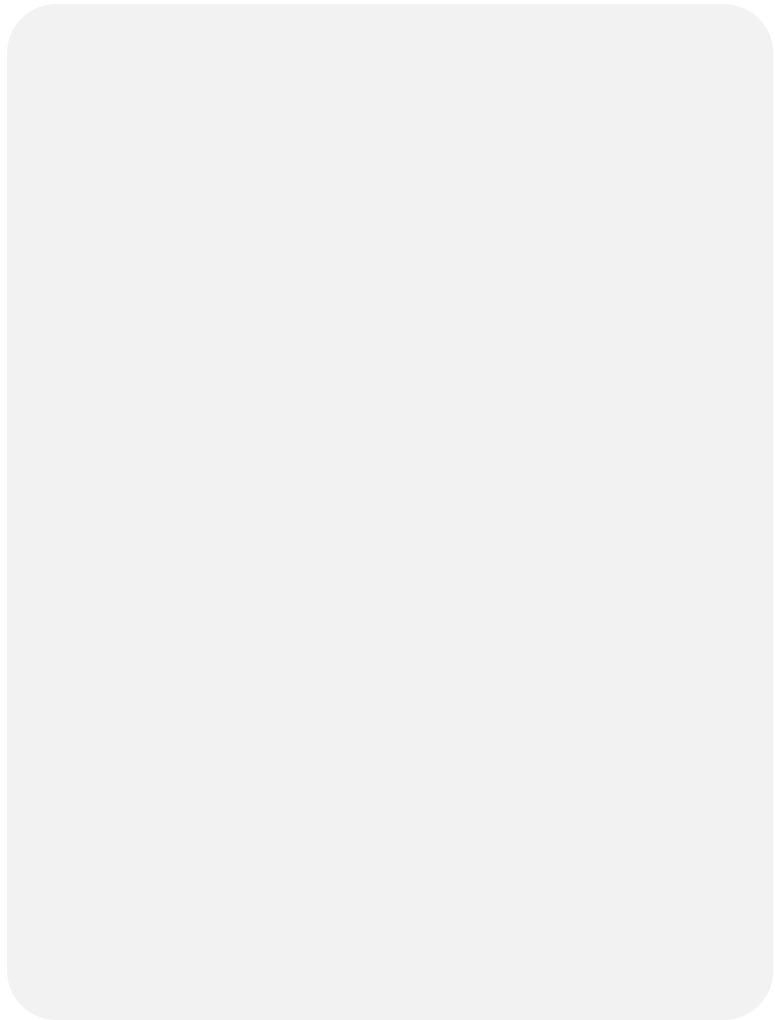


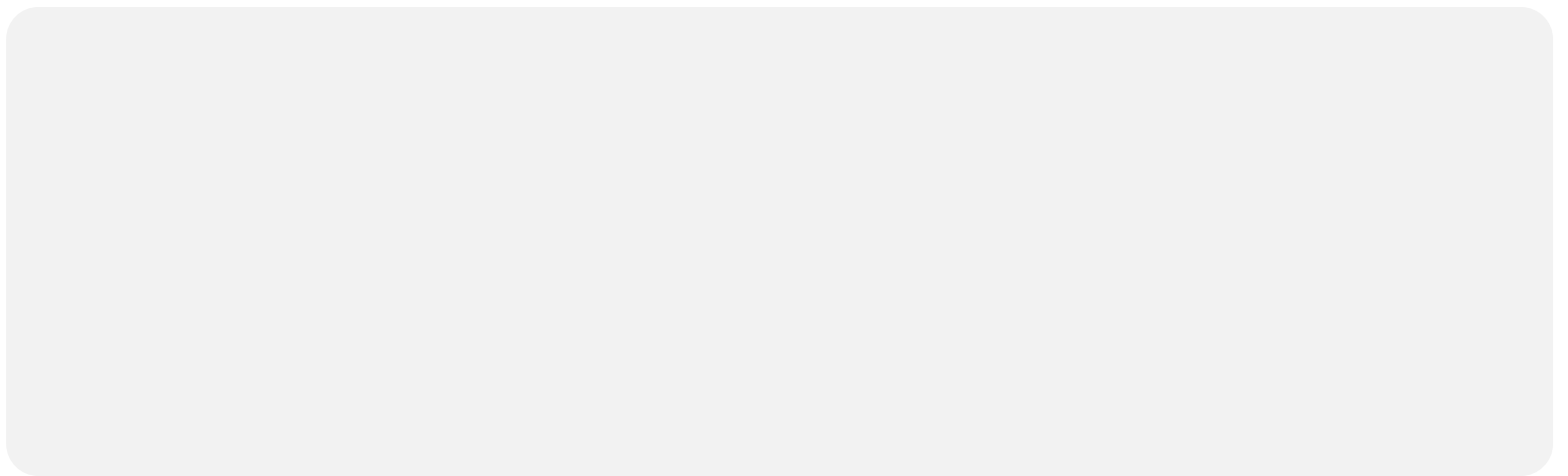
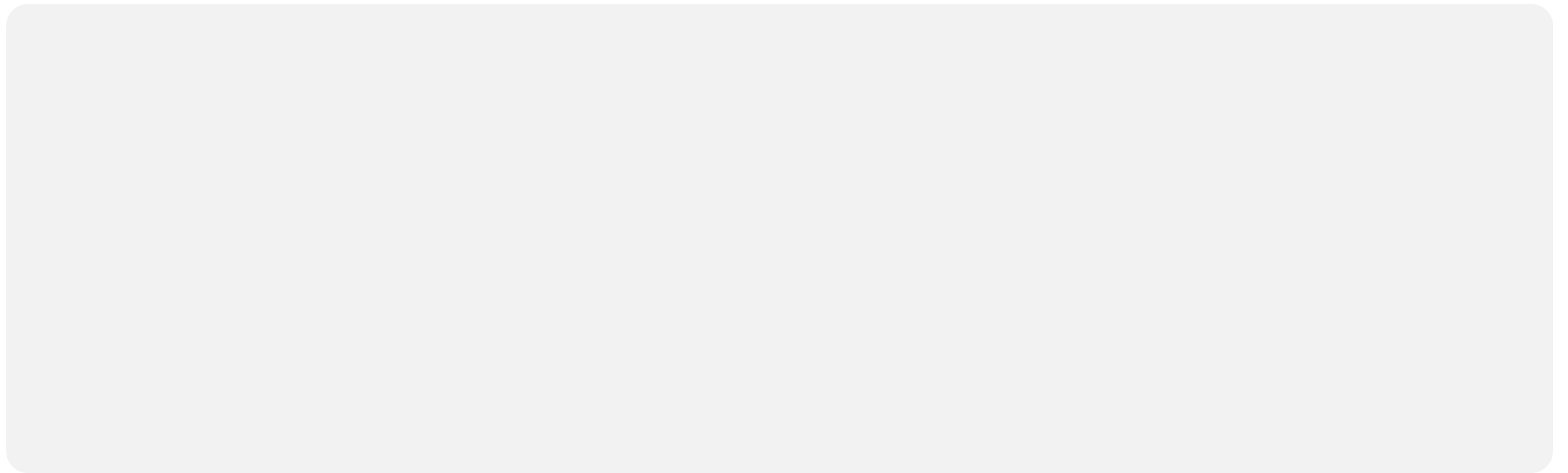
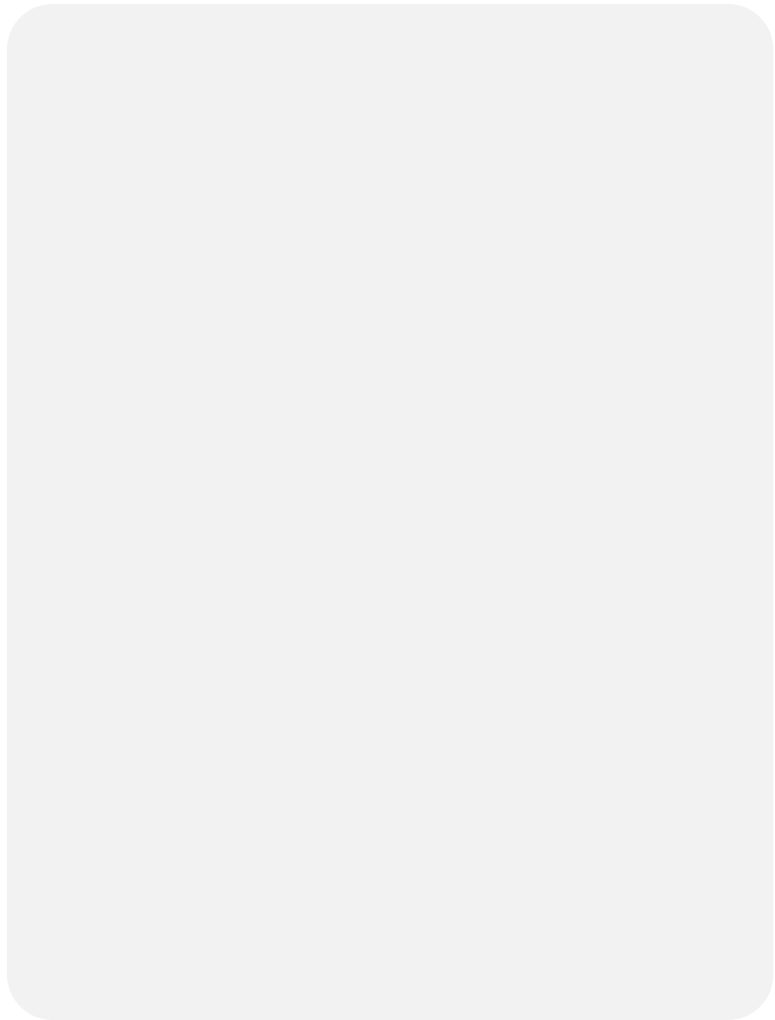


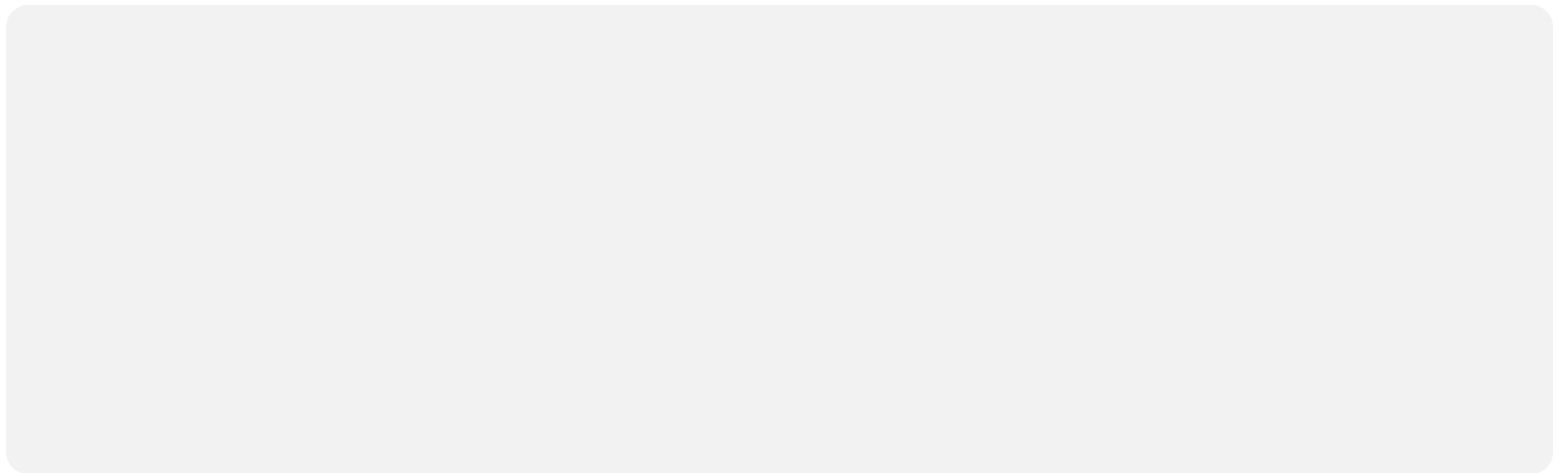
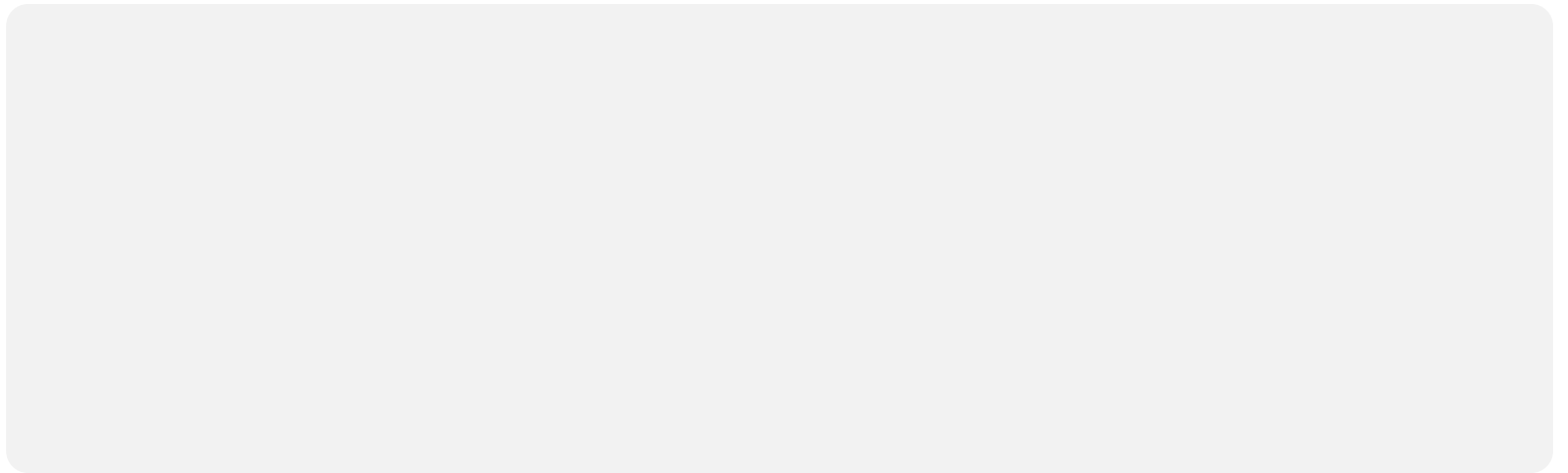
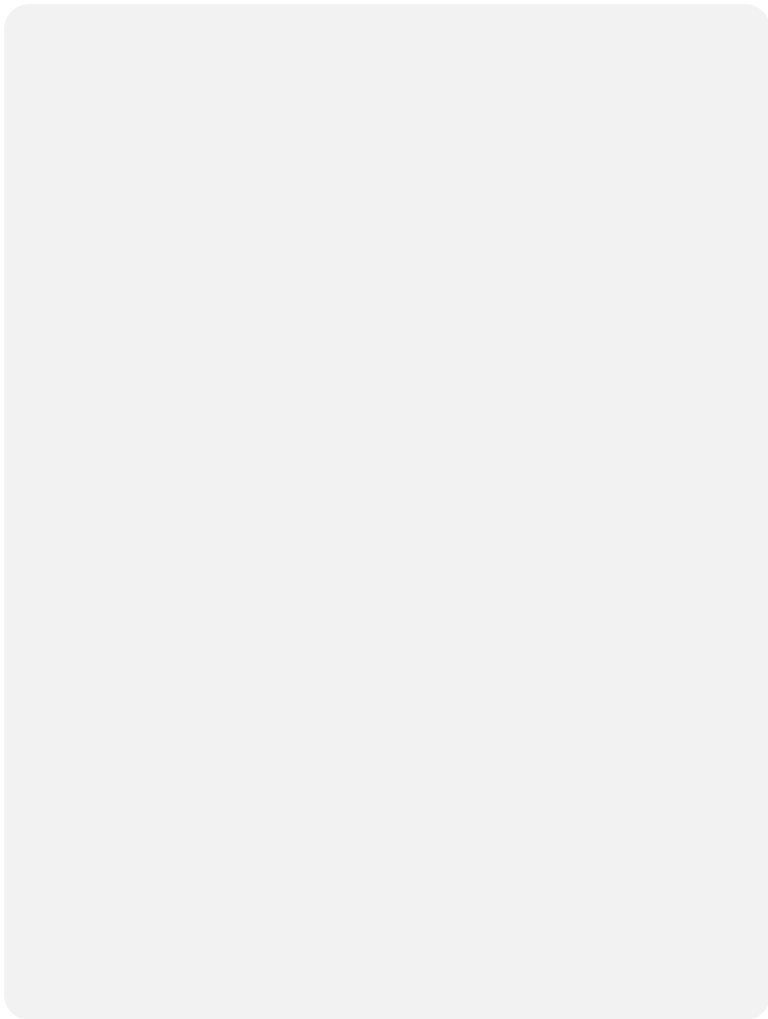


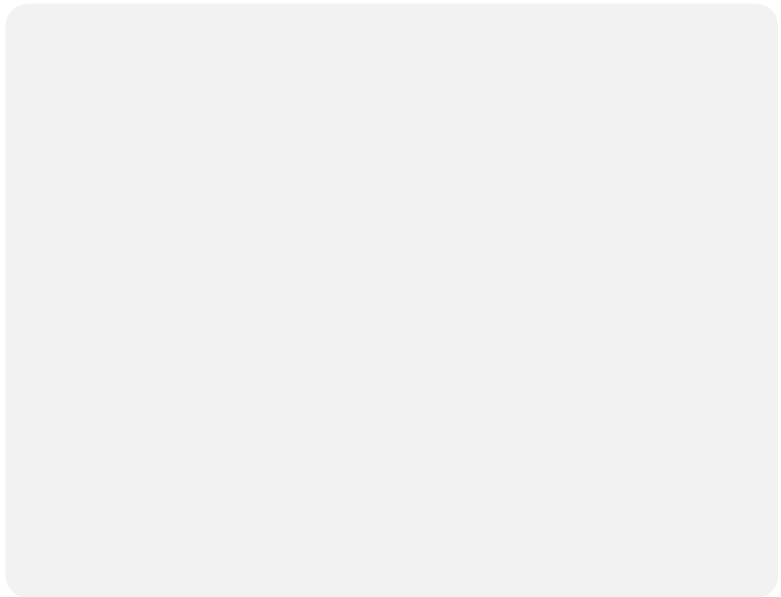
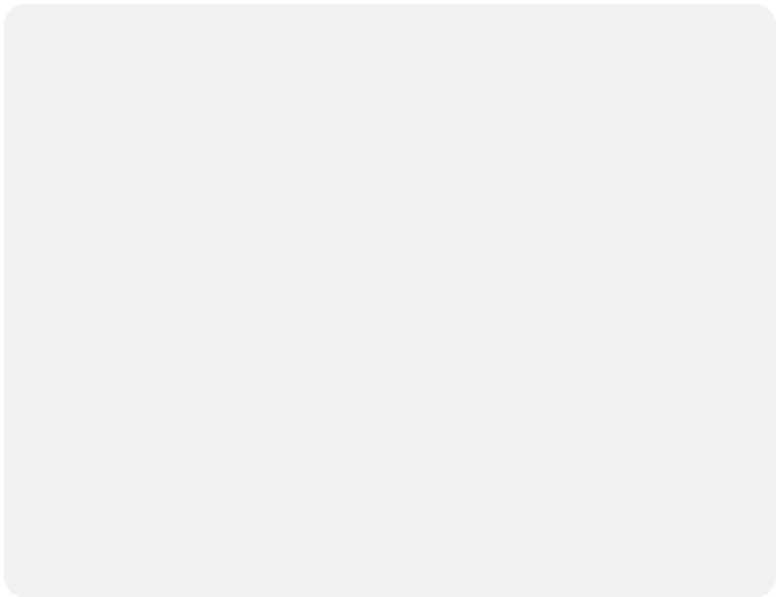
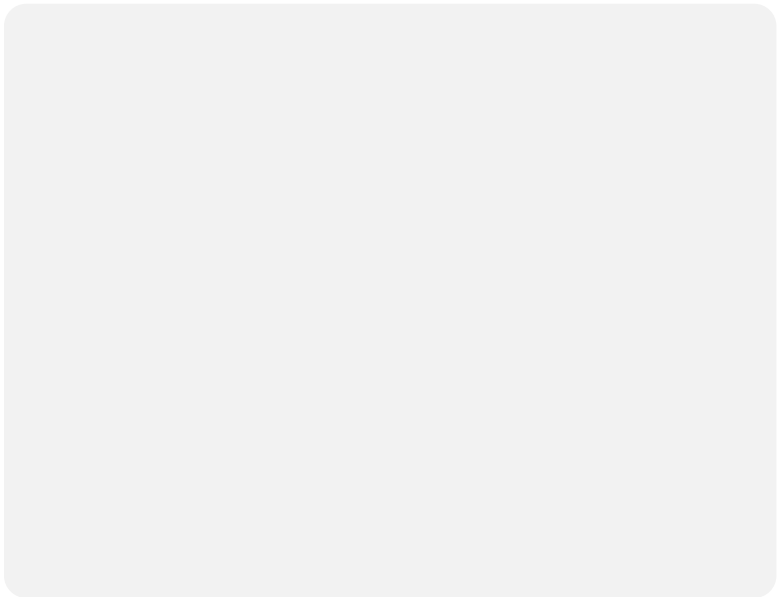
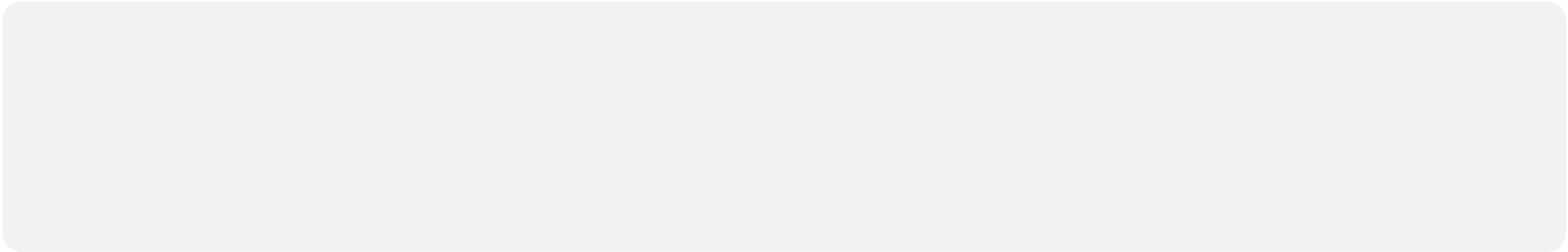


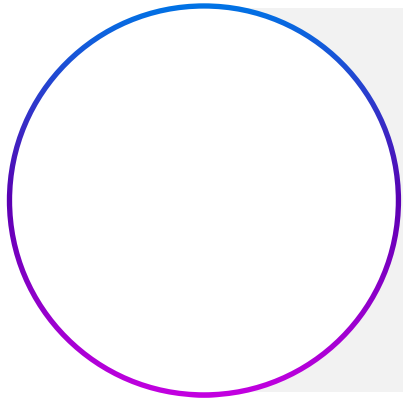


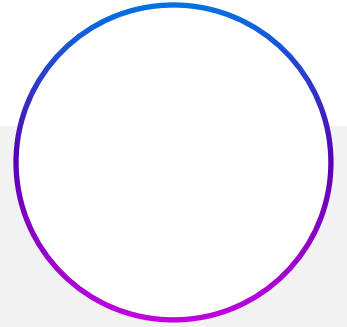
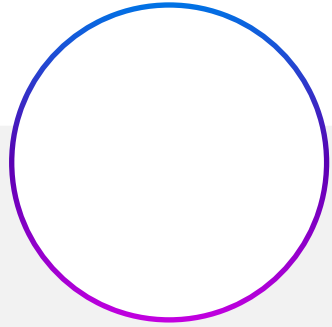
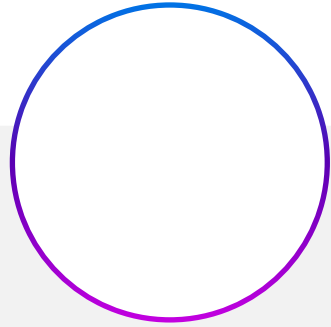


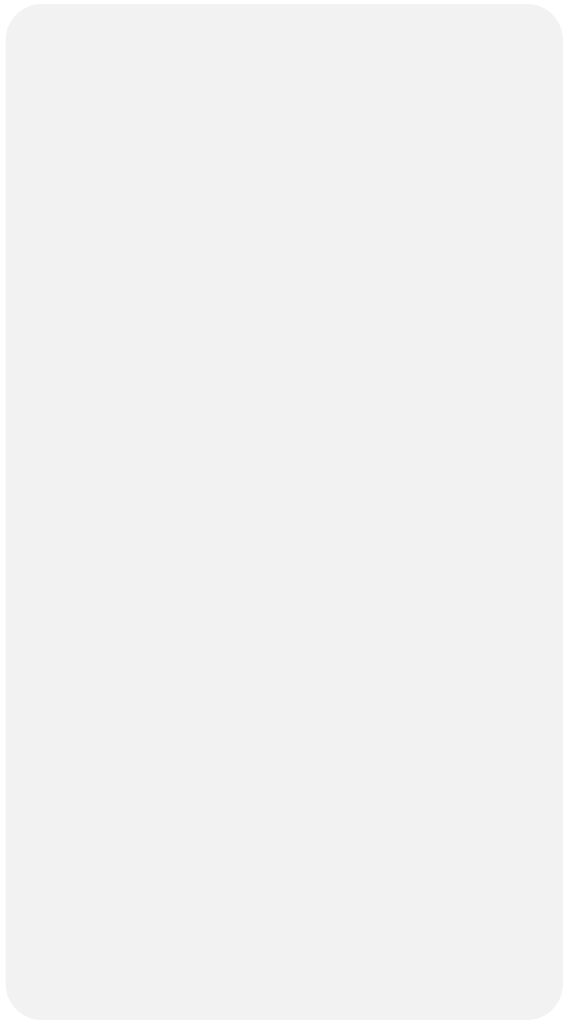
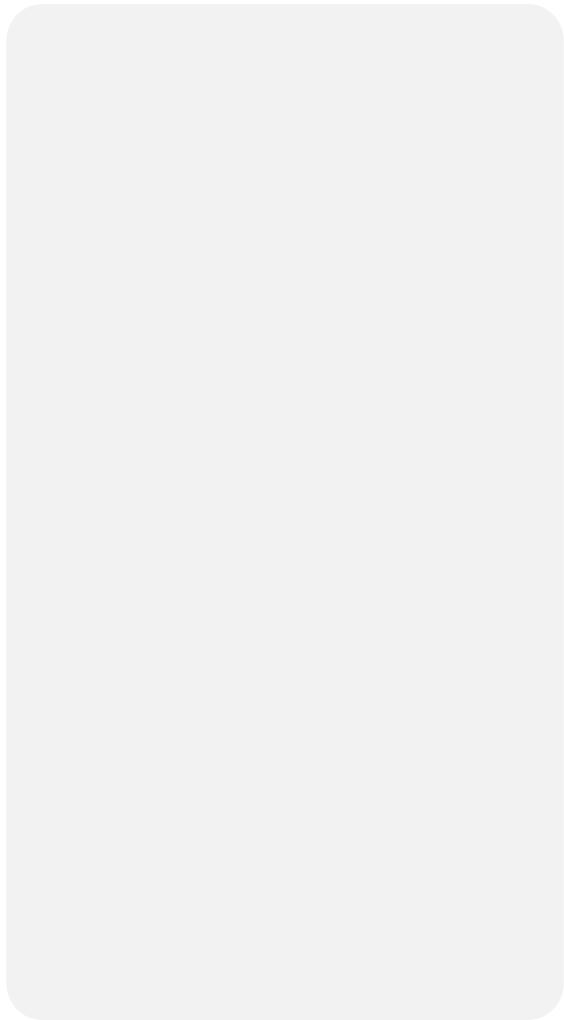
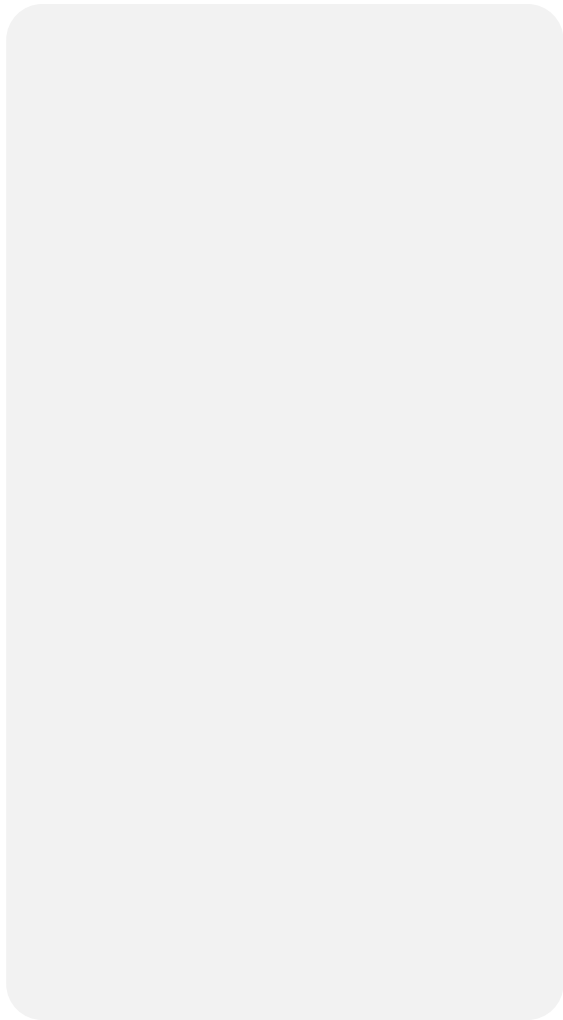
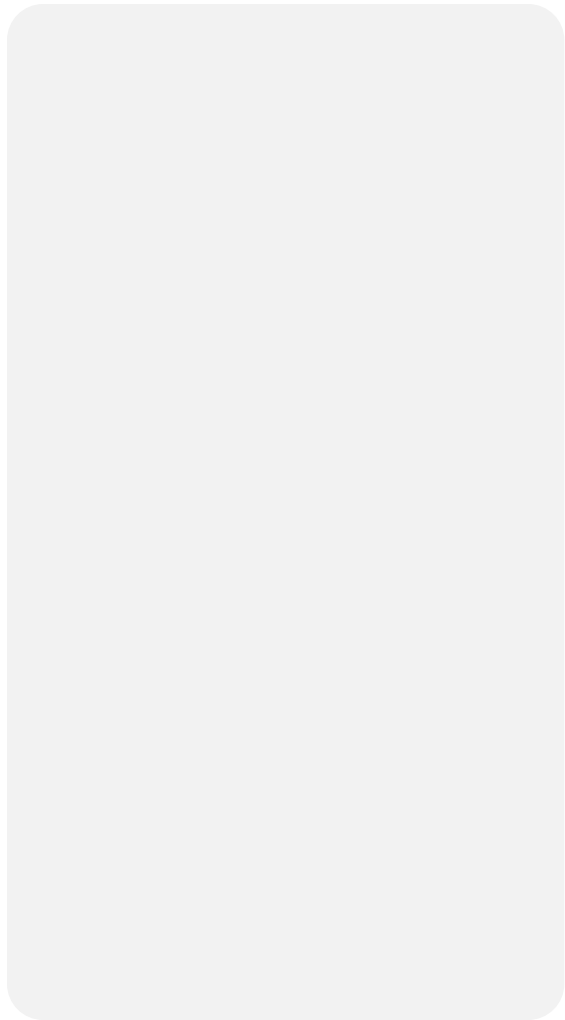


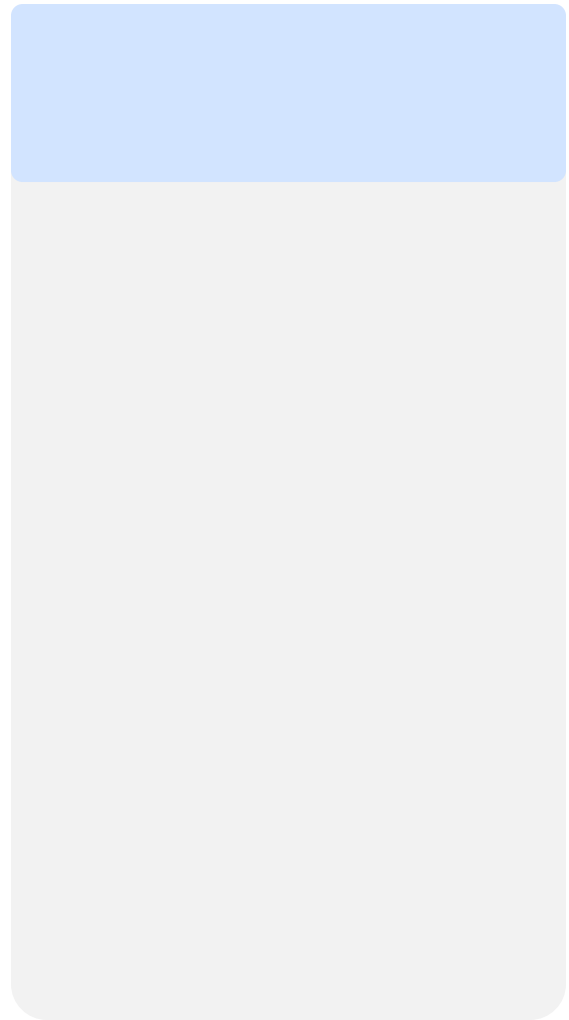
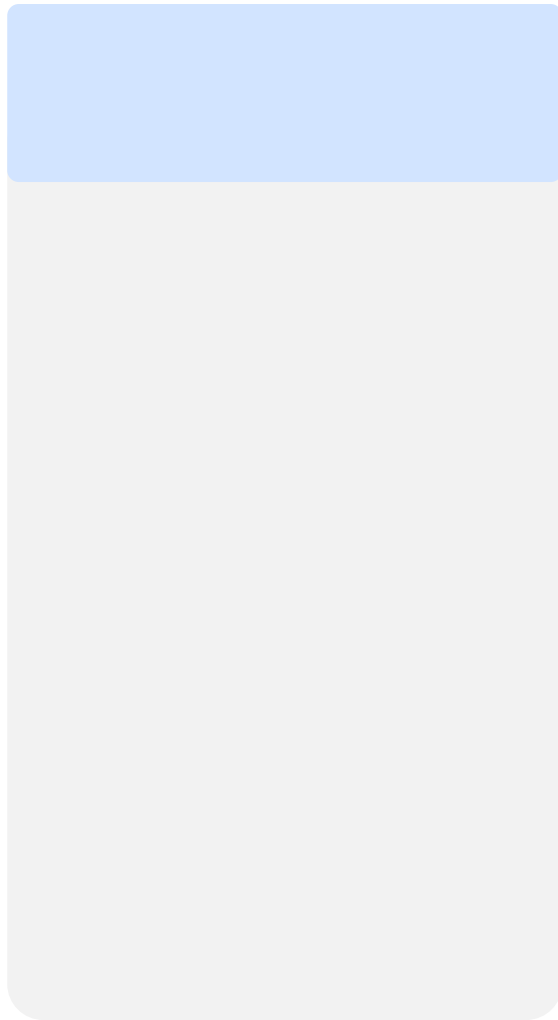
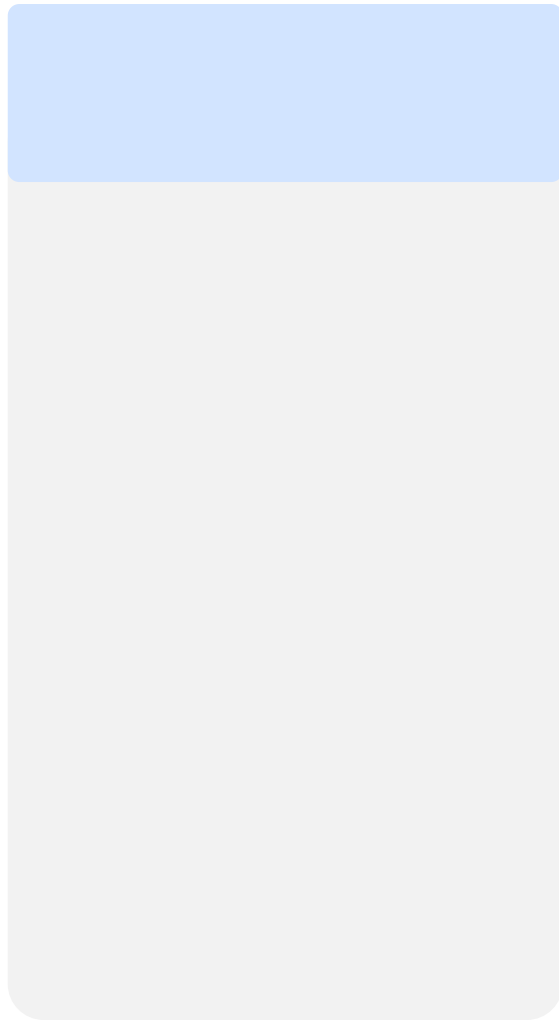
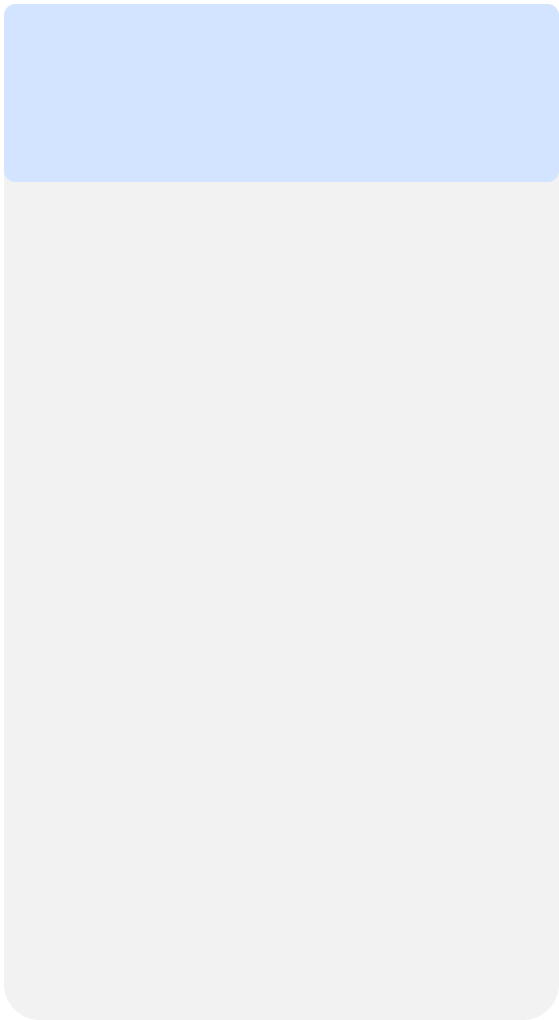


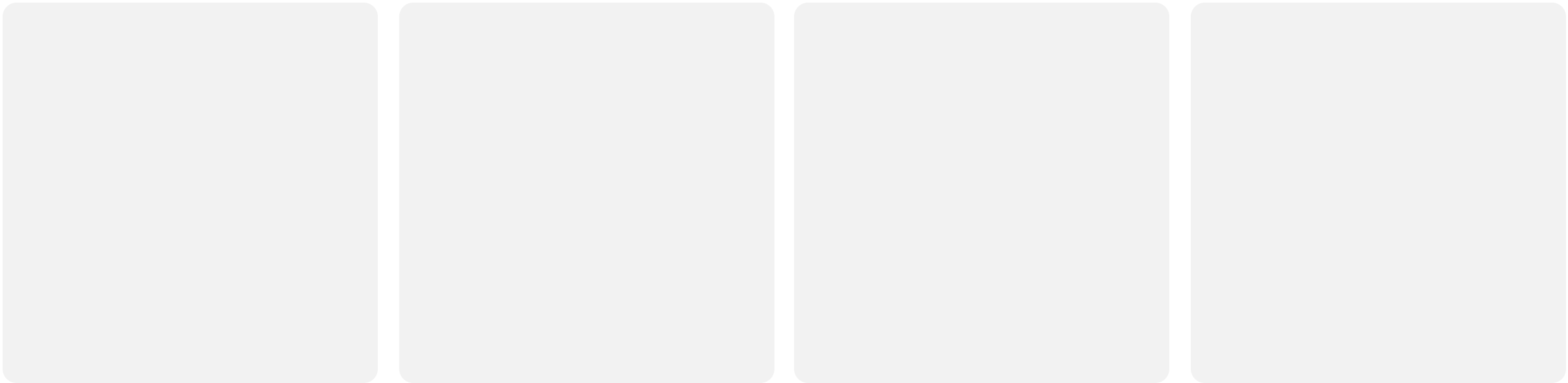




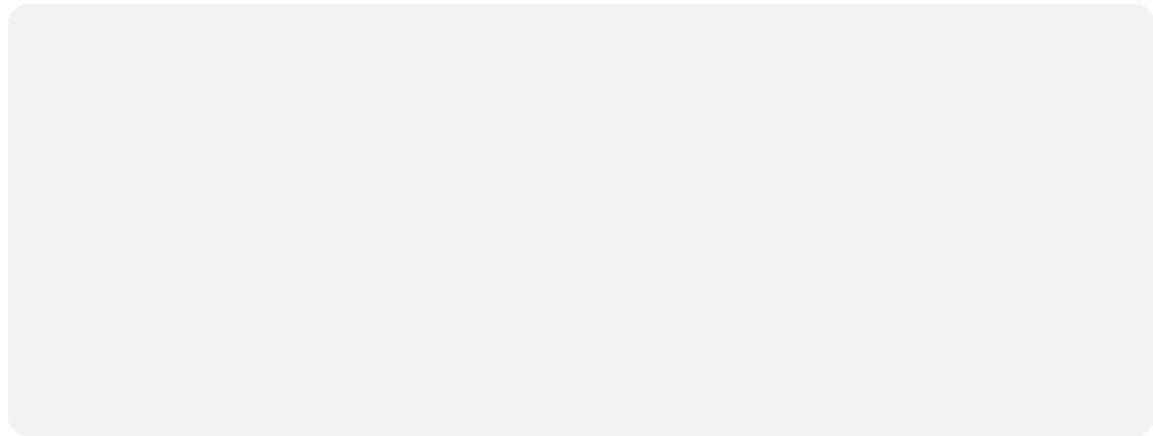
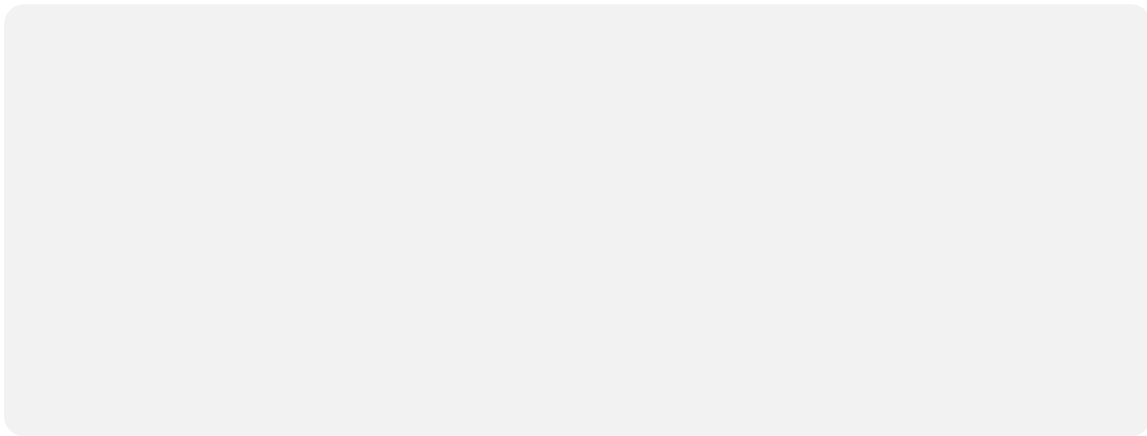
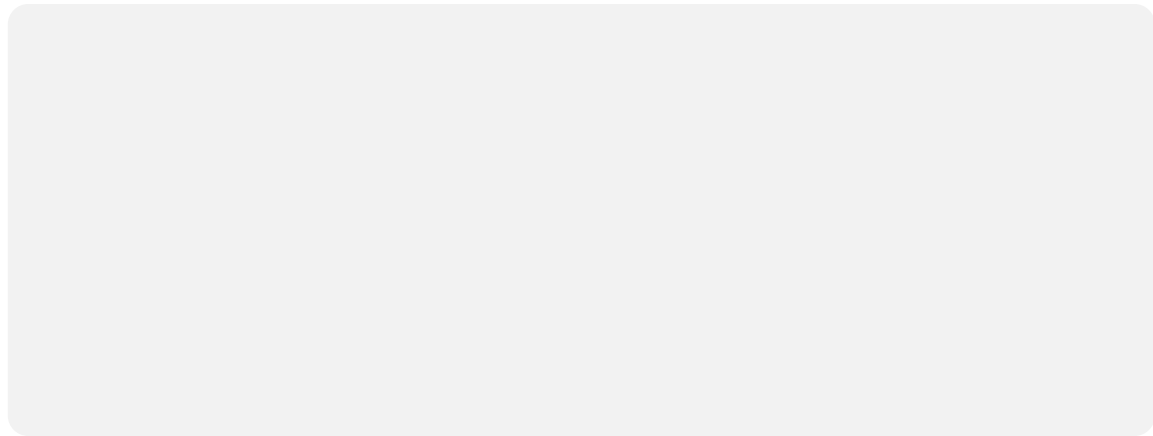
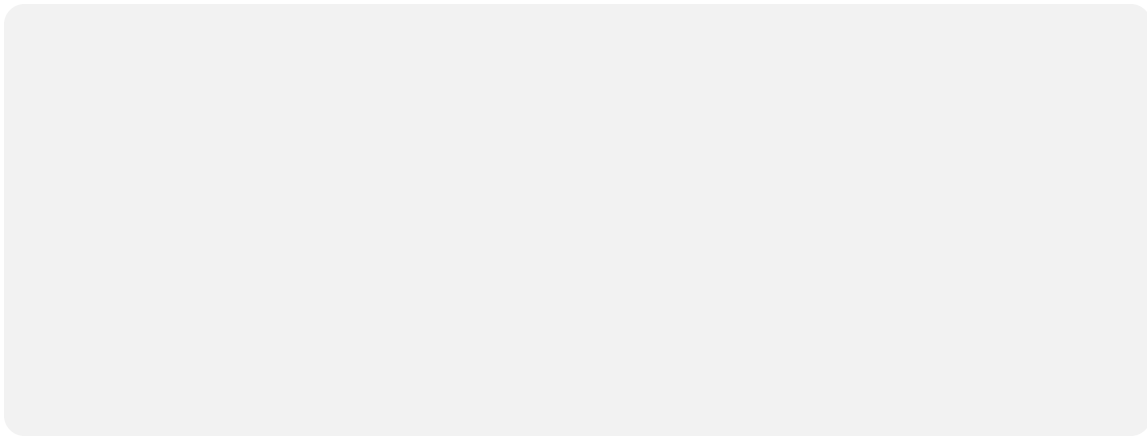




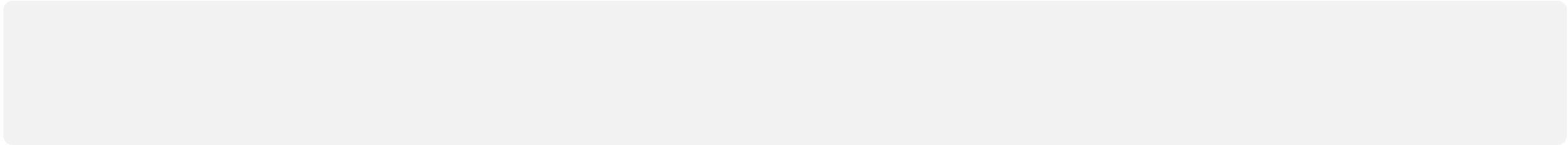
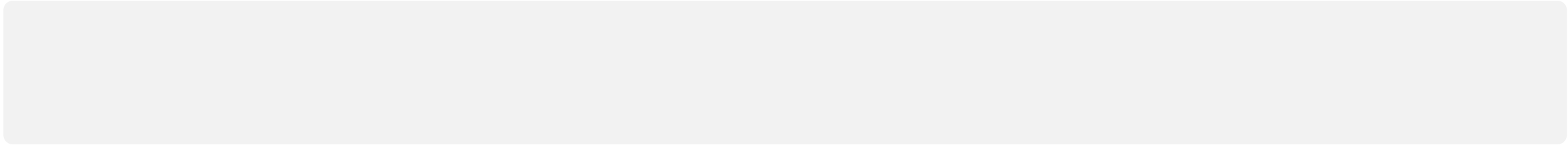
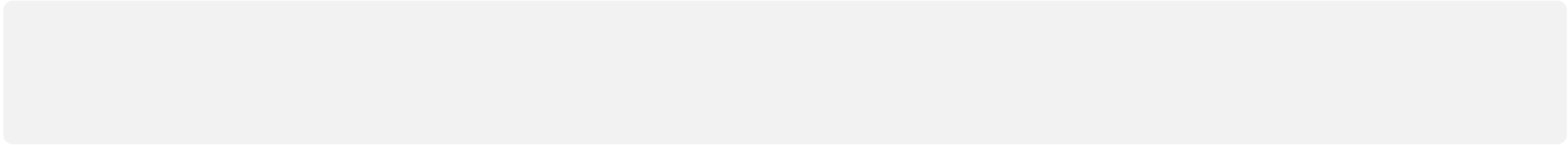
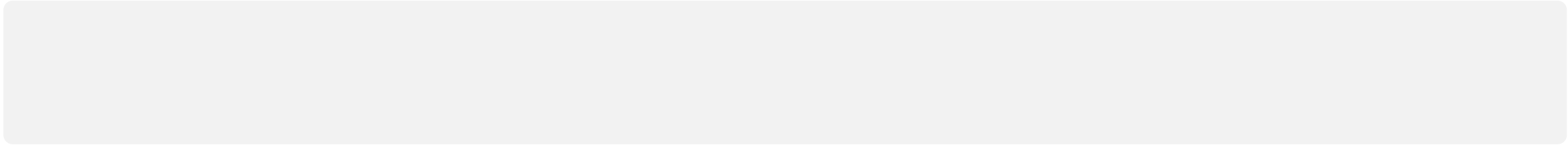


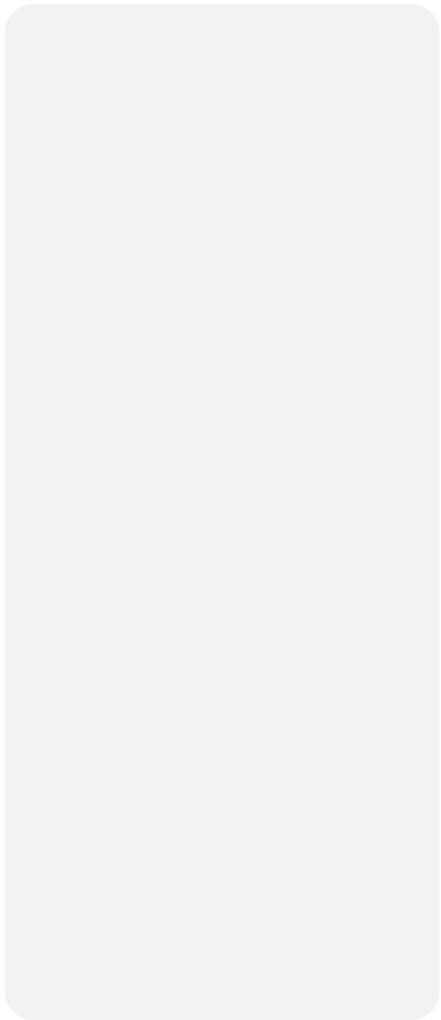
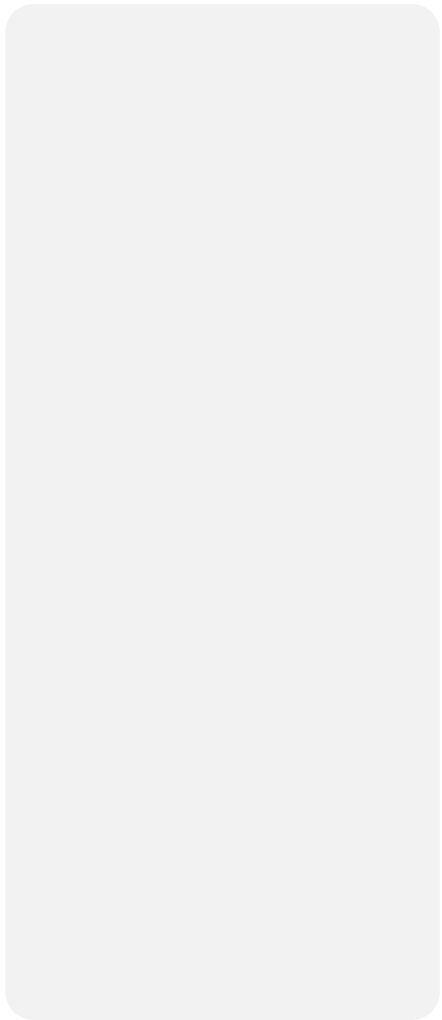
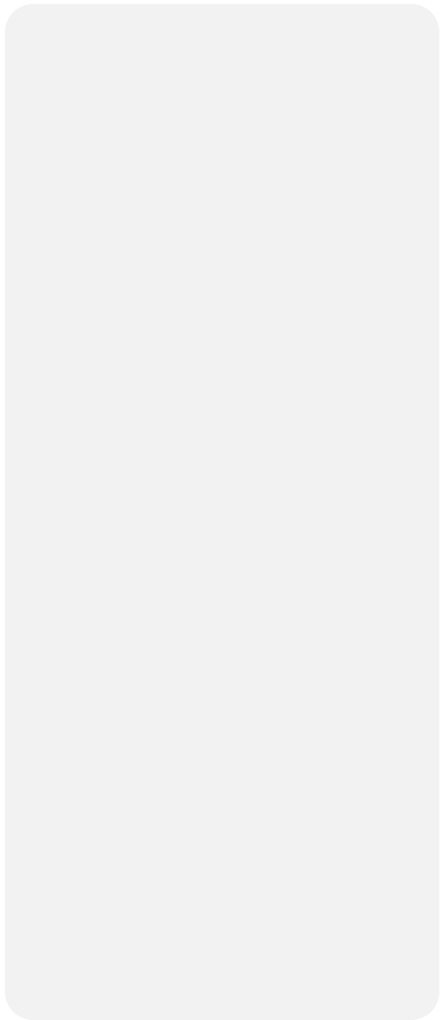
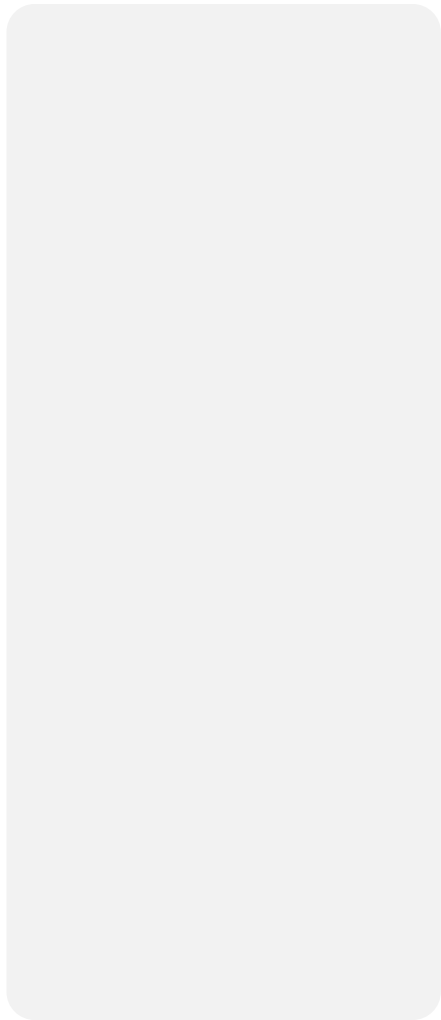
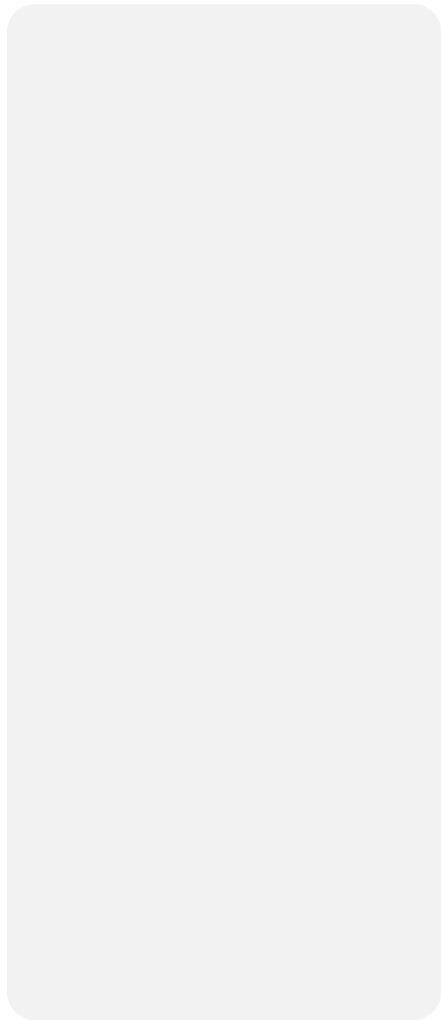


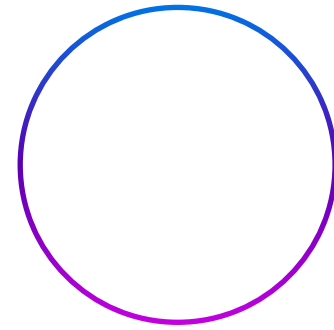
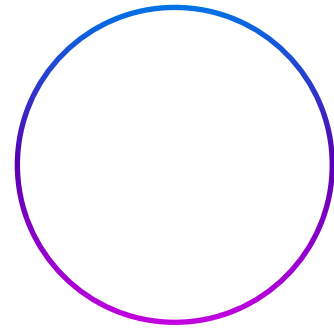
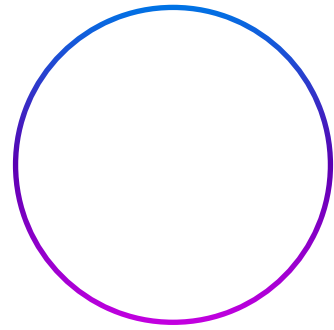
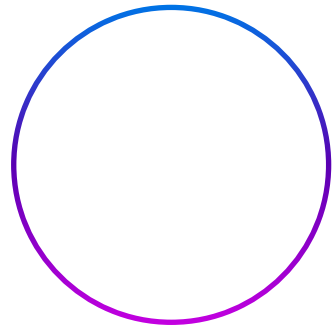
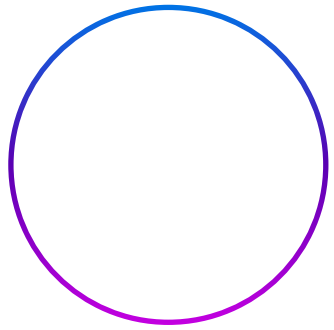






















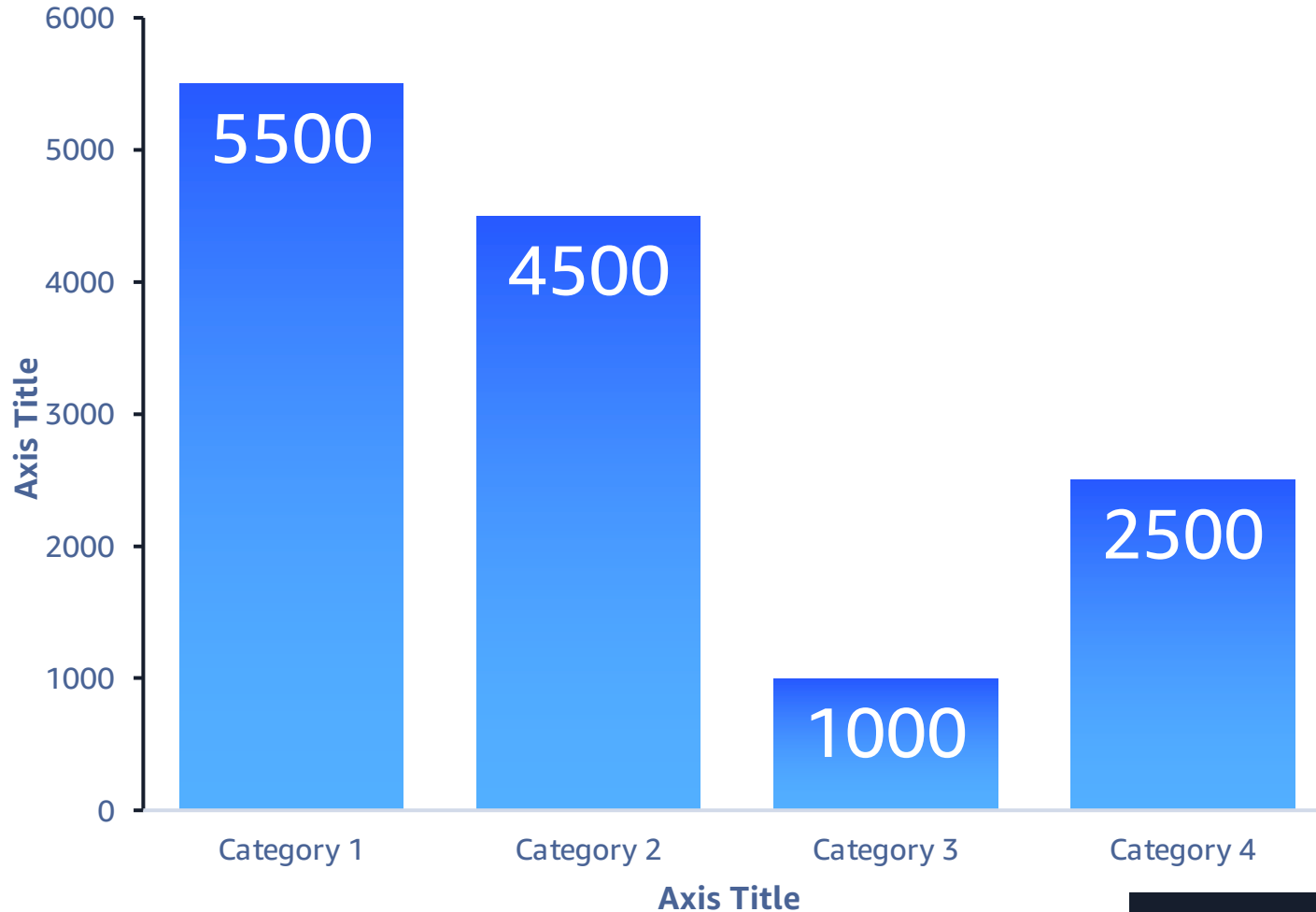




# Comparison

## To update the chart:

Select the graph then choose the menu option on the far right of the tool bar at the top "Chart Design" > "Edit Data". Bar height, spacing, and some axis labels will automatically adjust. See more info [in this Microsoft article](#).



## Main point

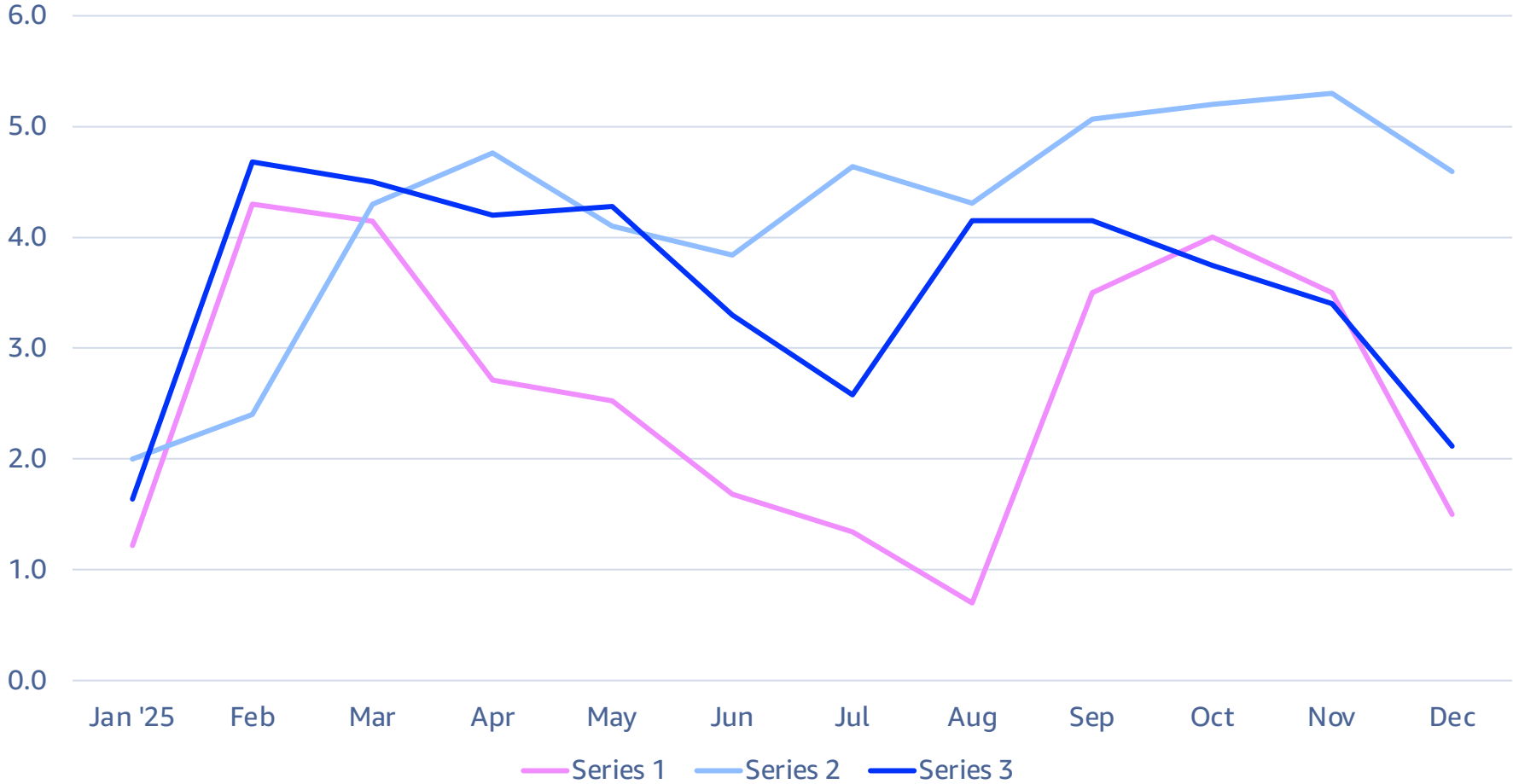
Full-sentence description of relationship between data points. Description may be multiple sentences in length.

237%

Note: You will have to decrease the font size of the above percentage if you are going over three digits.

# Trends

**To update the chart:**  
Select the graph then choose the menu option on the far right of the tool bar at the top "Chart Design" > "Edit Data". See more info [in this Microsoft article](#).



**Main point**  
Full-sentence description of data points.



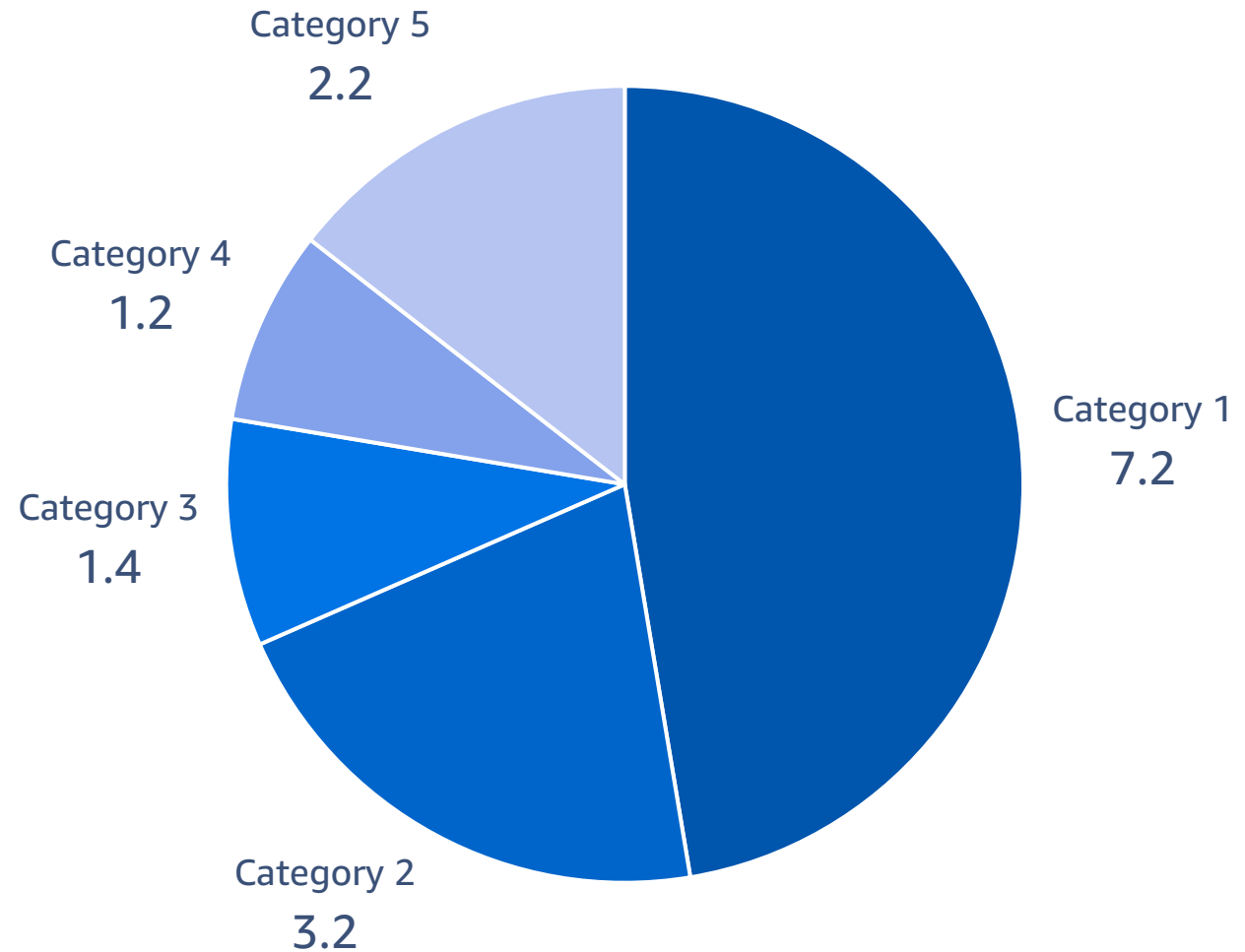
# Parts of a Whole

## To update the chart:

Select the pie graph, then choose the menu option on the far right of the tool bar "Chart Design" > "Edit Data". To change the color, double click the slice you want to change which will bring up a side bar on the right. Click the paint bucket > color > select what color you want to update it to. We recommend staying in the blue palette. The size, proportion, and labels of the slices will automatically adjust. See more info [in this Microsoft article](#).

## Main point

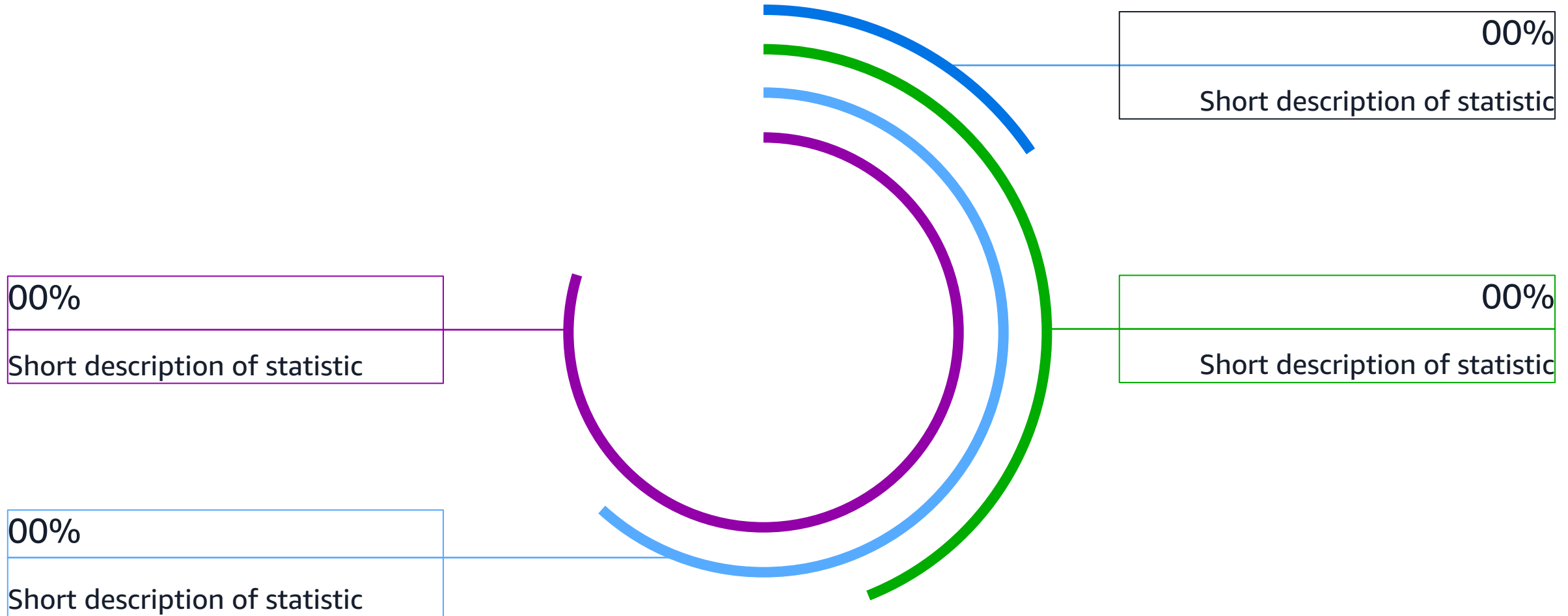
Full-sentence description of data points.



# Parts of a Whole

## To change the percentage depicted:

Select an arc and drag the yellow dot to make the arc longer or shorter (do NOT resize with the white handles/dots). Use visual judgement to approximate the percentage. **Note:** arcs should get longer as they approach the center (i.e., pink depicts a greater percentage than orange). You may need to reposition the data labels (click label, drag to reposition) to make sure they are connected to their corresponding arcs.

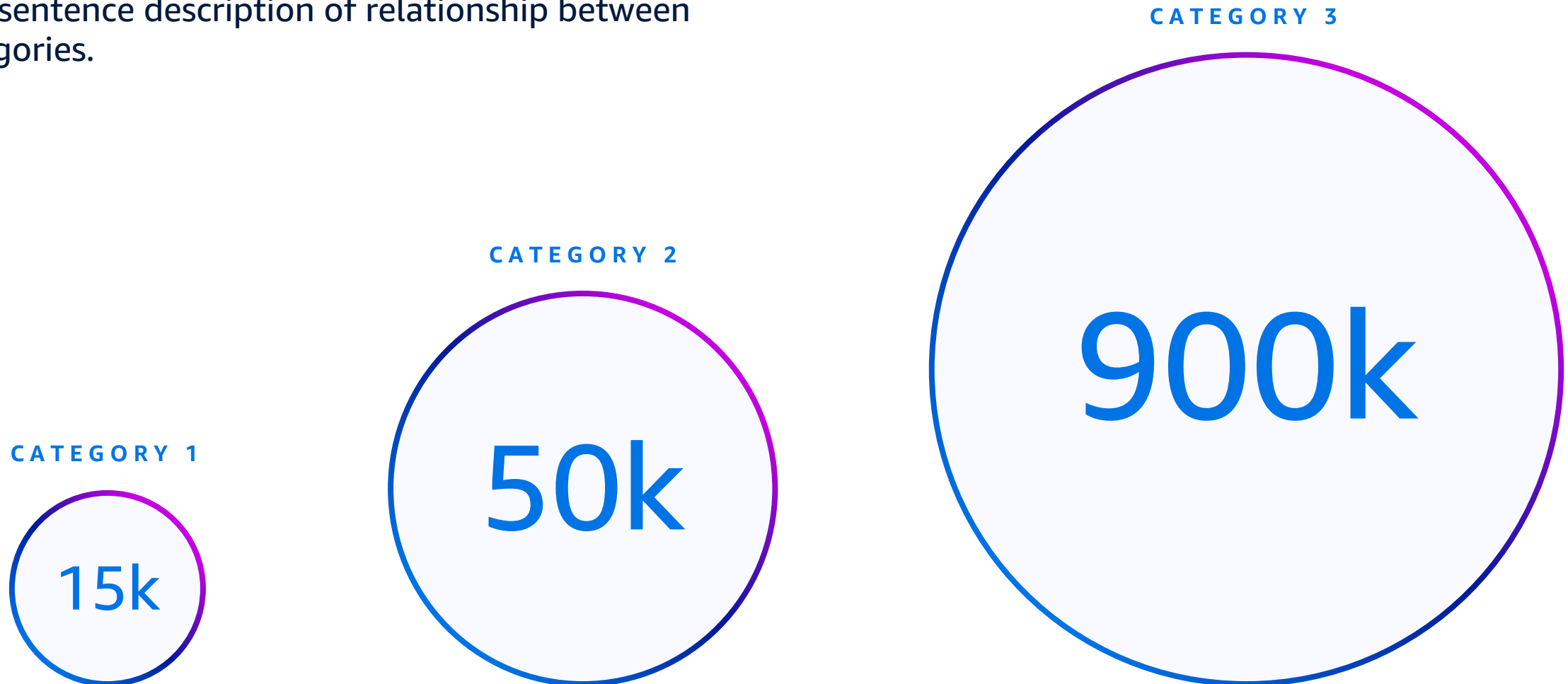


# Comparisons

To center the category label above the data circle:

To [keep the "Category" labels centered](#), hold shift and select the circle you want to adjust and its corresponding category label. Then choose the menu option on the far right of the tool bar "Shape Format" > "Align" > "Align Center". To align all three circles on the same baseline: Hold shift and select all three circle. Then choose the menu option on the far right of the tool bar "Shape Format" > "Align" > "Align Bottom".

Full-sentence description of relationship between categories.



# Comparisons

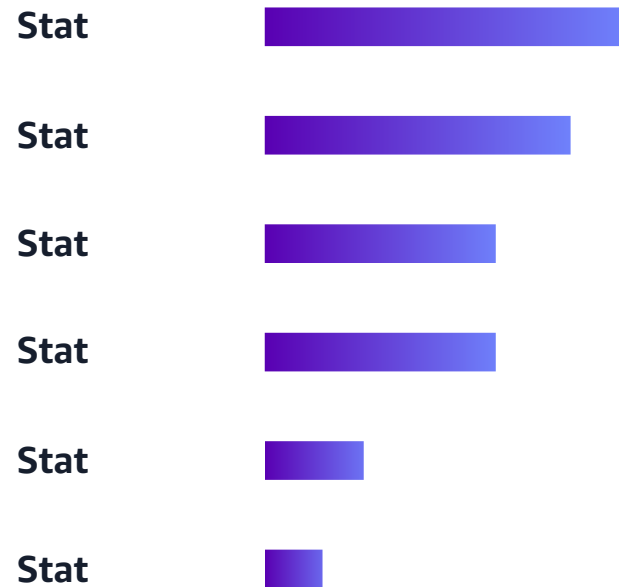
Full-sentence description of relationship between categories.

To resize bars:

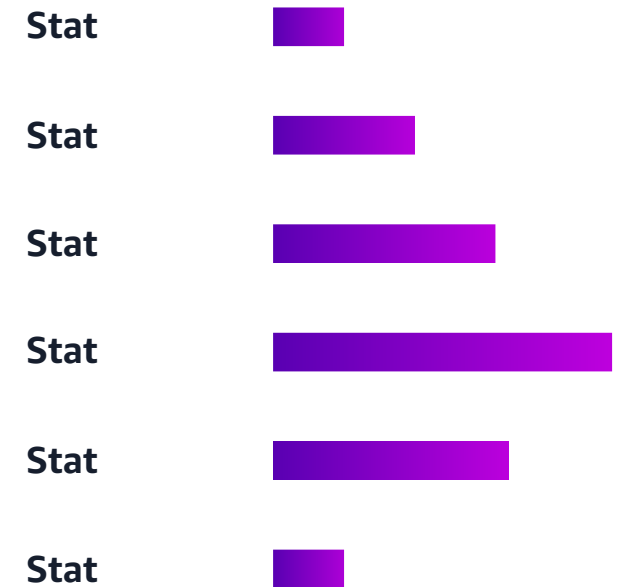
Click on one of the horizontal bars, hold shift, select the white dot, and drag to be the approximate relative proportion.

Note: holding down "Shift" when resizing keeps them flat horizontally.

## Category 1



## Category 2



# Comparisons

## How this graph works:

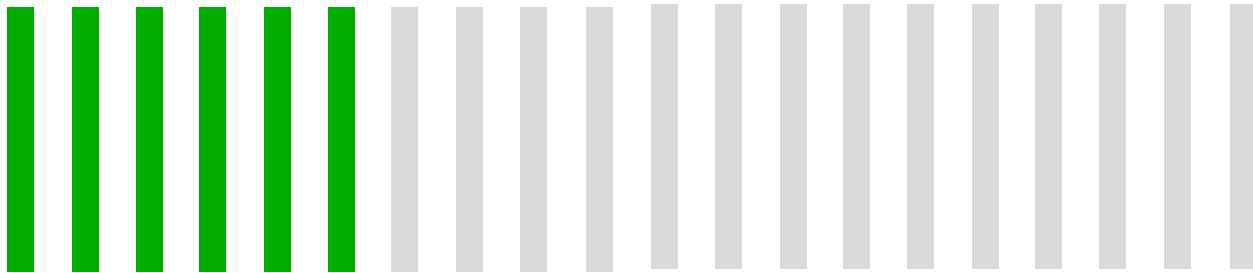
Both rows have 20 bars, each representing 5%. Round to the nearest 5% of your stat and fill in the appropriate number of bars with the orange or gold color.

## How to recolor bars:

Select the bar you want colored (if you want to color multiple, hold shift and click on all the bars you want to recolor) then choose the menu option on the far right of the tool bar "Shape Format" > "[Shape Fill](#)" [paint bucket](#), select appropriate color by selecting "Eyedropper" and select a point on one of the colored lines to get the same hue.



00% statistic description



00% statistic description

# Maps

## To resize, reposition, or duplicate circles:

Resize: Select circle and drag on white handles/dots to adjust size.

Reposition: Select circle and drag to new place on the map.

Duplicate: Click circle and then copy / paste. Click on newly pasted circle and reposition as needed. Make sure to only select the circle and not the map image when moving.

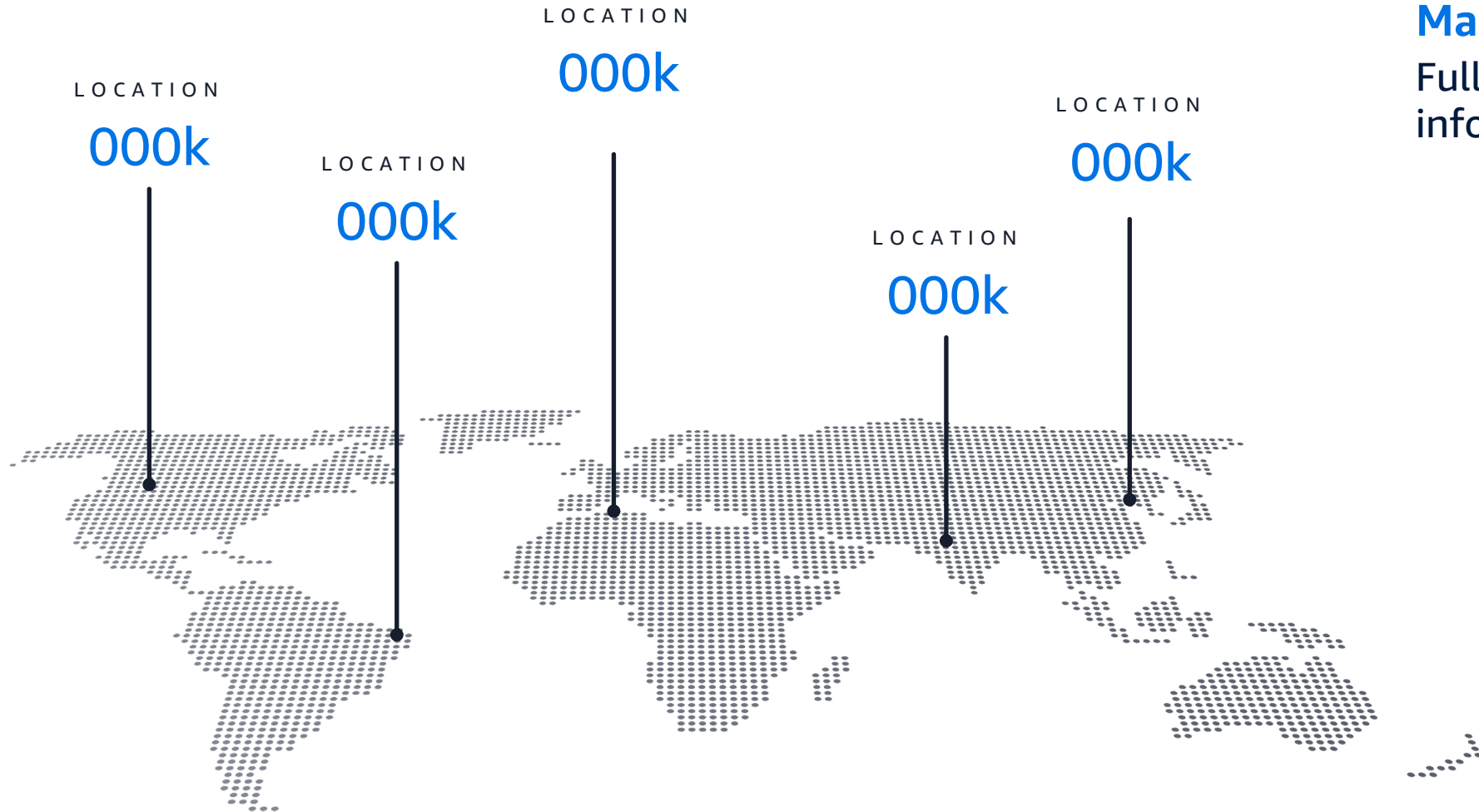
Do not move or recolor map.



# Maps

To update text and move markers as needed:

Hold shift and click on the white line and the text box above the line to move them together.  
The map is somewhat abstracted, so location need not be precise.



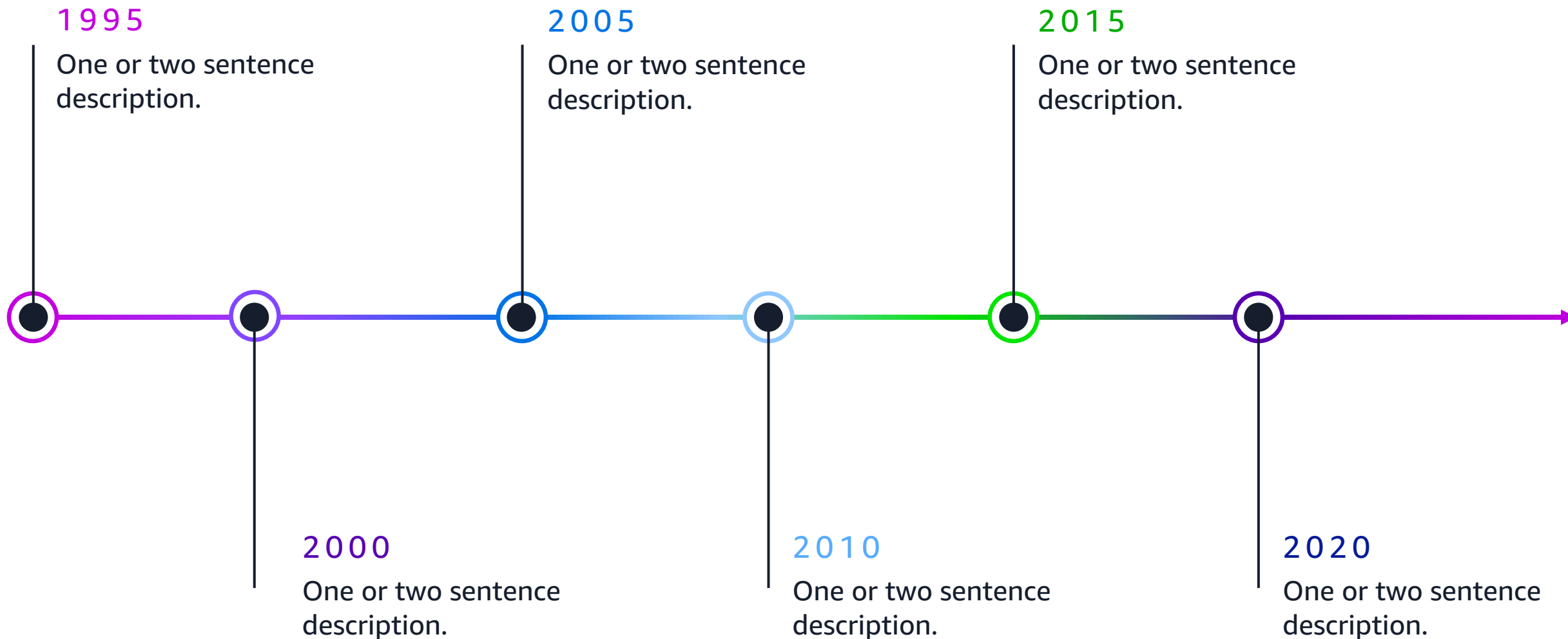
## Main point

Full-sentence description of information conveyed by map.

# Journey

## To shift date points:

Hold shift and select one of the white lines (which has a white dot with it) and the corresponding outline circle (colored) and then drag to a new position. To recolor the outline circle to match the new placement on the gradient line, select only the outline circle, choose the menu option on the far right of the tool bar "Shape Format" > "Shape Outline" > "Eyedropper" and select a point on the gradient line near the circle to match the color.



# Journey

## Have overlapping events?

To suggest a journey step to continue past a later step, you can adjust the length of the bars.

## To adjust the length of the bars:

Select the rectangle shape, click and drag white dot on the right to extend the length.

Journey step

Journey step

Journey step

Journey step

Journey step

# Typographic

## To update data numbers:

Select number text box and type in relevant data stat.

Depending on how many numbers you need to depict for each stat, you might need to decrease font size to fit.

00

Longer, full-sentence  
description of number  
or statistic

00%

One/two-word  
description

00M

Longer, full-sentence  
description of number or  
statistic

# Typographic

To update data number:

Select number text box and type in relevant data stat.

00k

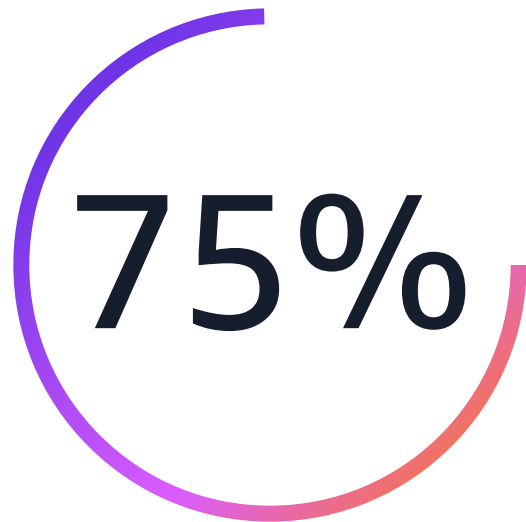
Description

# Typographic

To change the percentage depicted:

Select one of the gradient arcs and drag the yellow dot to make the arc longer or shorter (do NOT resize with the white handles/dot).

Use visual judgement to approximate the percentage.



Description of  
number or statistic



Description of  
number or statistic