



Please Sit With Your Group

L-6

Unsupervised Learning and Reinforcement Learning I

MAIB7002

Fundamentals of AI Models and Technologies

Tutorial

If you struggle with coding, attend a tutorial

Course code walkthroughs and Q&A on existing material

17:00–18:30 · Hosts: Neel & Jack

Date	Venue
Friday, 9 October	A205 1 & 2
Tuesday, 13 October	A205 1 & 2

These replace the tutorials on Wednesday, 7 October and Wednesday, 14 October

Exam

Saturday, 17 October · 18:30–20:30

Cyberport 4, EFG

Notes Allowed

Two A4 sheets

Printed or handwritten notes on both sides

Calculators

Not permitted

Question Format

Question types include MCQs

Coding: change parameters in existing code

No writing code from scratch

Mock Exam Paper

Released Saturday, 10 October

Similar format and difficulty to the exam

Assignment 2

Released: Can We Trust This Model?

Due Tuesday, 13 October

Individual Work · 10% of the Course

Submit only [a2.ipynb](#) on Moodle

Group Project

Submit by Wednesday, 14 October
Exhibition on Friday, 16 October

20% of the Course · One Submission per Group on Moodle

Project	Docker project + notebook or report
Poster	A1-size poster PDF
Each Member	Contribution and AI-use statement

Poster Exhibition and Peer Reviews

Friday, 16 October · CP-LT104, Cyberport 4

Session	Round I	Round II
1	14:00–15:00	15:30–16:30
2	18:30–19:30	20:00–21:00

Attend your assigned session

Present in one round; review in the other

Follow your named allocation in the project brief

Explore both sessions and learn from all projects

A Problem and Solution Worth Sharing

Choose an interesting problem that matters

New to the Class

Familiar Method

Explain how it works

Provide new insights

Both are welcome

Teach through an example or demo

What will your classmates learn?

Your Project in Two Sentences

Summarise your current project in two sentences

Problem

We aim to address [problem]
because [why it matters
and to whom]

Solution

We propose to use
[method or approach]
to [how it addresses the problem]

Use these as prompts; you can phrase your sentences differently

One submission per group on Ed

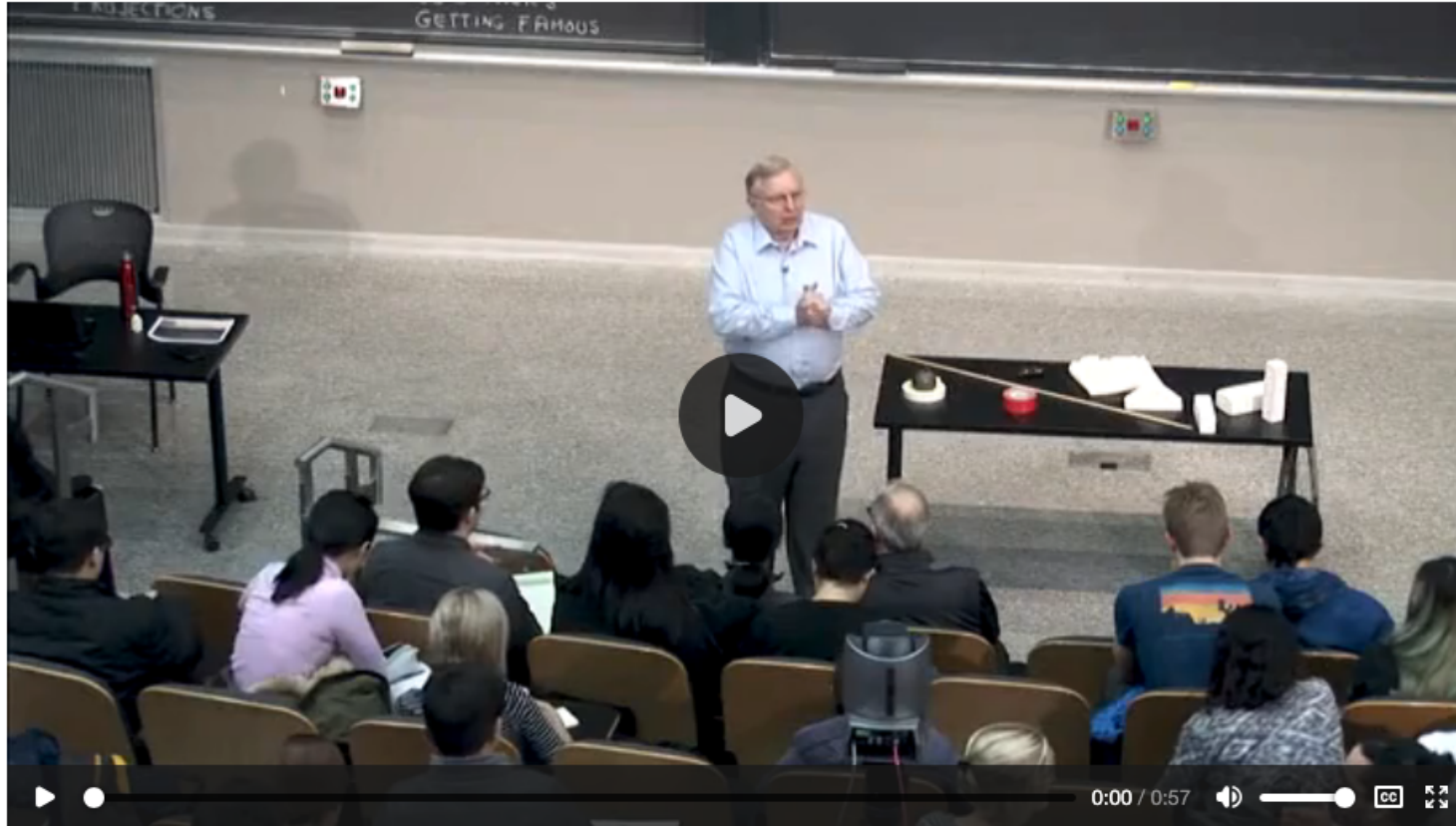
What Will Your Classmates Learn?

What is one specific thing your classmates should learn from your project?

Which example, result or demo will help them understand it?

One submission per group on Ed

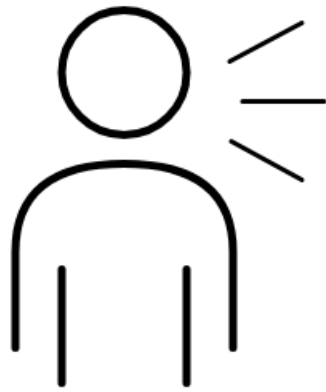
Why Communication Matters



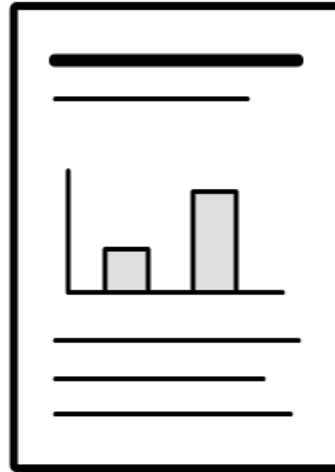
Patrick Henry Winston · MIT · How to Speak · Opening excerpt

Help others understand and value your ideas.

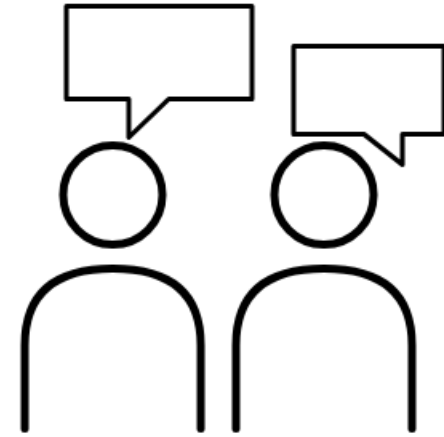
Why a Poster?



Explain the Main Idea



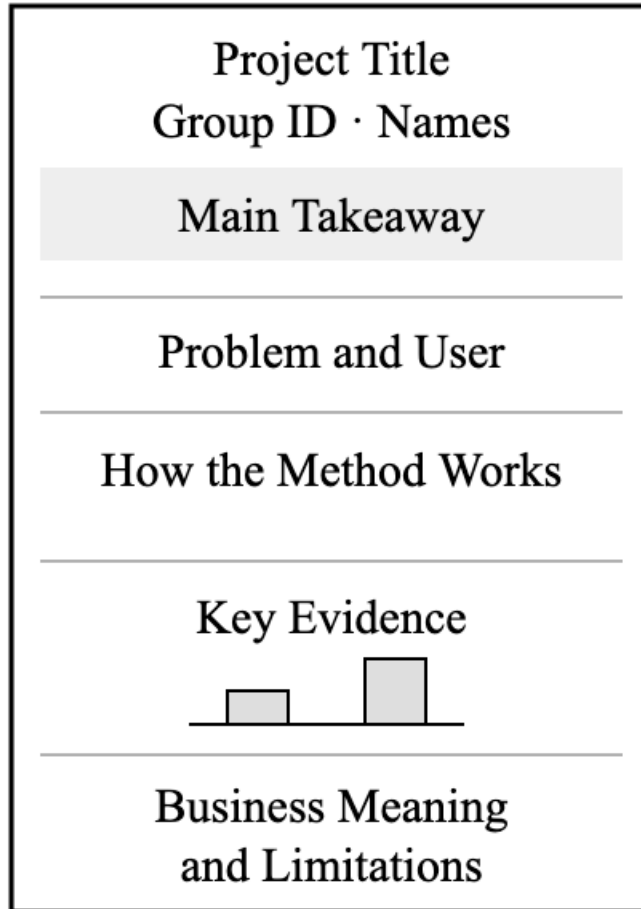
Show
the Evidence



Respond to Questions

Help Classmates Learn From Your Work

Make Your Poster Clear



One A1 Page · Portrait

594 mm wide × 841 mm tall

Make One Takeaway Prominent

What should the visitor remember?

Let Figures Explain the Work

A worked example and a clear comparison

Use Short Text and Readable Labels

Guide the reader from problem to conclusion



PIGS IN SPACE: EFFECT OF ZERO GRAVITY AND AD LIBITUM FEEDING ON WEIGHT GAIN IN CAVIA PORCELLUS

SPACEEXES

Colin B. Purrington

6673 College Avenue, Swarthmore, PA 19081 USA

ABSTRACT:

One ignored benefit of space travel is a potential elimination of obesity, a chronic problem for a growing majority in many parts of the world. In theory, when an individual is in a condition of zero gravity, weight is eliminated. Indeed, in space one could conceivably follow ad libitum feeding and never even gain an gram, and the only side effect would be the need to upgrade one's stretchy pants("exercise pants"). But because many diet schemes start as very good theories only to be found to be rather harmful, we tested our predictions with a long-term experiment in a colony of Guinea pigs (*Cavia porcellus*) maintained on the International Space Station. Individuals were housed separately and given unlimited amounts of high-calorie food pellets. Fresh fruits and vegetables were not available in space so were not offered. Every 30 days, each Guinea pig was weighed. After 5 years, we found that individuals, on average, weighed nothing. In addition to weighing nothing, no weight appeared to be gained over the duration of the protocol. If space continues to be gravity-free, and we believe that assumption is sound, we believe that sending the overweight — and those at risk for overweight — to space would be a lasting cure.

INTRODUCTION:

The current obesity epidemic started in the early 1960s with the invention and proliferation of elastane and related stretchy fibers, which released wearers from the rigid constraints of clothes and permitted monthly weight gain without the need to buy new outfits. Indeed, exercise today for hundreds of million people involve only the act of wearing stretchy pants in public, presumably because the constrictive pressure forces fat molecules to adopt a more compact tertiary structure (Xavier 1965).

Luckily, at the same time that fabrics became stretchy, the race to the moon between the United States and Russia yielded a useful fact: gravity in outer space is minimal to nonexistent. When gravity is zero, objects cease to have weight. Indeed, early astronauts and cosmonauts had to secure themselves to their ships with seat belts and sticky boots. The potential application to weight loss was noted immediately, but at the time travel to space was prohibitively expensive and thus the issue was not seriously pursued. Now, however, multiple companies are developing cheap extra-orbital travel options for normal consumers, and potential travelers are also creating new ways to pay for products and services that they cannot actually afford. Together, these factors open the possibility that moving to space could cure overweight syndrome quickly and permanently for a large number of humans.

We studied this potential by following weight gain in Guinea pigs, known on Earth as fond of ad libitum feeding. Guinea pigs were long envisioned to be the "Guinea pigs" of space research, too, so they seemed like the obvious choice. Studies on humans are of course desirable, but we feel this current study will be critical in acquiring the attention of granting agencies.

MATERIALS AND METHODS:

One hundred male and one hundred female Guinea pigs (*Cavia porcellus*) were transported to the International Space Laboratory in 2010. Each pig was housed separately and deprived of exercise wheels and fresh fruits and vegetables for 48 months. Each month, pigs were individually weighed by duct-taping them to an electronic balance sensitive to 0.0001 grams. Back on Earth, an identical cohort was similarly maintained and weighed. Data was analyzed by statistics.

RESULTS:

Mean weight of pigs in space was 0.0000 ± 0.0002 g. Some individuals weighed less than zero, some more, but these variations were due to reaction to the duct tape, we believe, which caused them to be alarmed push briefly against the force plate in the balance. Individuals on the Earth, the control cohort, gained about 240 g/month ($p = 0.0002$). Males and females gained a similar amount of weight on Earth (no main effect of sex), and size at any point during the study was related to starting size (which was used as a covariate in the ANCOVA). Both Earth and space pigs developed substantial dewlaps (double chins) and were lethargic at the conclusion of the study.

CONCLUSIONS:

Our view that weight and weight gain would be zero in space was confirmed. Although we have not replicated this experiment on larger animals or primates, we are confident that our result would be mirrored in other model organisms. We are currently in the process of obtaining necessary human trial permissions, and should have our planned experiment initiated within 80 years, pending expedited review by local and Federal IRBs.

ACKNOWLEDGEMENTS:

I am grateful for generous support from the National Research Foundation, Black Hole Diet Plans, and the High Fructose Sugar Association. Transport flights were funded by SPACE-EXES, the consortium of wives divorced from insanely wealthy space-flight startups. I am also grateful for comments on early drafts by Mañana Athletic Club, Corpus Christi, USA. Finally, sincere thanks to the Cuy Foundation for generously donating animal care after the conclusion of the study.

LITERATURE CITED:

NASA. 1982. Project STS-XX: Guinea Pigs. Leaked internal memo.
Sekulic, S.R., D. D. Lukač, and N. M. Naumović. 2005. The Fetus Cannot Exercise Like An Astronaut: Gravity Loading Is Necessary For The Physiological Development During Second Half Of Pregnancy. *Medical Hypotheses*. 64:221-228
Xavier, M. 1965. Elastane Purchases Accelerate Weight Gain In Case-control Study. *Journal of Obesity*. 2:23-40.



CONSTRUCTION & SIMULATION ANALYSIS OF AN IMPROVED ACTIN FILAMENT MODEL

Here, we present a new model of the actin filament (F-actin) that incorporates the global structure of a recently published model¹ but also conserves internal stereochemistry. The improved quality of the model is apparent in a comparison made of the model with other recent F-actin models using molecular dynamics (MD) simulation, monitoring a number of structural determinants.

In addition, simulations of the model are carried out in states with both ADP or ATP bound and local hydrogen-bonding differences characterized. The results point to the significance of a direct interaction of Glu17 with ATP for activation of ATPase activity after the G-to-F-actin transition.

Jeremy C. Smith

IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY



Thomas J. DeLamater

IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY

Kenneth C. Holmes

IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY

Frank Nij

IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY

INTRODUCTION

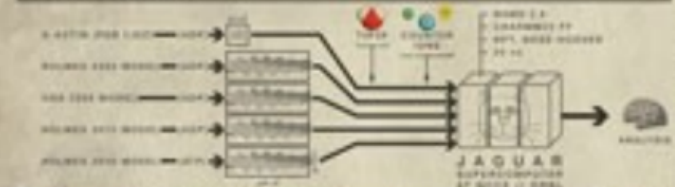
The atomic-detail structure of F-actin is still unknown. We propose a new model of F-actin, "Holmes 2010", which was built using a straightforward approach in which priority was given to keeping the stereochemistry within the actin protomer intact while altering the position of the two actin domains to account for the global conformational change during the G-to-F-actin transition. A comparison is made of the structures and dynamics of Holmes 2010 with Oda 2009¹ and Holmes 2004² by subjecting them to MD simulation.

The nucleotide effects (ADP or ATP) on the conformation of the filament are also studied with a particular focus on the G-to-F-actin ATPase activation.

HOLMES 2010 MODEL

- Starting structure: G-actin X-ray structure PDB (J6Z)³
- Global conformational change during G-to-F-actin transition
- Separate fitting of two domains to Oda model
- DNase I binding loop: replace α -helix with Oda coordinates
- Phalloidin added and coordinates refined against fiber diffraction data⁴ and EM data⁵ weighted in favor of the fiber data
- Final radius of gyration: 23.7 Å

METHODS



REFERENCES

- Oda, T., et al. (2009) The structure of the actin filament in the presence of phalloidin. *Science* 325, 1132-1135.
- Holmes, K.C., et al. (2004) The structure of the actin filament in the presence of phalloidin. *Science* 306, 2055-2058.
- Winkler, F., et al. (2007) The structure of the actin filament in the presence of phalloidin. *Science* 316, 1503-1506.
- Winkler, F., et al. (2007) The structure of the actin filament in the presence of phalloidin. *Science* 316, 1503-1506.
- Winkler, F., et al. (2007) The structure of the actin filament in the presence of phalloidin. *Science* 316, 1503-1506.

RESULTS

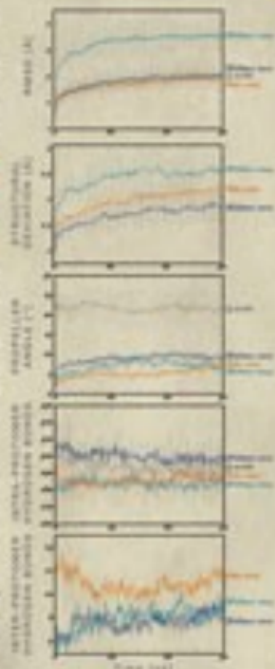
The new "Holmes 2010" model was developed and MD simulations of 30 ns length were performed to compare the structure and stability with those of Oda 2009 and Holmes 2004. A further simulation of the Holmes 2010 model was performed in the ATP state.



Most properties were assessed on the protomer level and then averaged over the 13 protomers in the filament.

The average RMSD of the Oda 2009 and Holmes 2010 model protomers is very close to that of the simulation of the G-actin crystal structure. Among the 3 tested models, the structural deviation of Holmes 2004 is the lowest and that of Holmes 2010 the highest. As expected, the propeller angle formed by the four subunits of the actin molecule is much higher in the G-actin simulation than in the simulations of the F-actin models.

The number of hydrogen bonds within the actin molecules is the highest in the MD of Holmes 2010, even higher than that of G-actin. But most hydrogen bonds between protomers are formed in the MD of Oda.



Among the three models, only Holmes 2004 has a DNase binding loop in α -helical conformation, which unfolds entirely in 6 of the 13 protomers. Further, all protomers in Holmes 2004 show unfolding of some of the secondary structural elements. In the simulations, the hydrophobic plug showed a high degree of structural variation in Holmes 2004 and some variation in the Oda 2009 MD but was highly conserved among the Holmes 2010 protomers. Residues 348-357 form an α -helix that remains stable in the Oda simulations but unfolds in some protomers of the two Holmes simulations. The WH2-binding domain (residues 165-170) forms a β -hairpin in the simulations of both Holmes models but a bend in the Oda MD. Residues 127-137 exhibit partial unfolding in most Oda protomers.



Previous studies suggest the significance of Glu17 for ATP hydrolysis.^{6,7} The twist of the actin monomer upon integration into the filament brings Glu17 close to ATP. In the MD Glu17 forms a stable H-bond with the oxygen of the β -phosphate group, and may thus play a direct role in hydrolysis.

CONCLUSIONS

The MD comparison reflects the evolution in quality of the actin models over the last years. The Holmes 2010 model possesses both a global conformation in agreement with the recent Oda model and a consistent intraprotomer stereochemistry. As such it should form a useful basis for atomic-detail investigations into F-actin structure-dynamics-function relationships.

ACKNOWLEDGEMENTS

We thank Isabella Daidone and Durba Sengupta for helpful discussion and acknowledge support from Deutsche Forschungsgemeinschaft and a DOE Research and Development "Systems Biology" Grant. Computations were performed on the Jaguar Cray XT4 supercomputer at the National Center for Computational Sciences at Oak Ridge National Lab.





I can barely read that!

Ooh, that's pretty!

Placeholder text blocks and a line graph.

Placeholder text blocks and a bar chart.

Placeholder text blocks and a scatter plot.

Placeholder text blocks and a horizontal bar chart.

ANIMATE YOUR SCIENCE.

Placeholder text blocks and a line graph.

Placeholder text blocks and a circular diagram.

Placeholder text blocks and a list of icons.

Placeholder text blocks and a horizontal bar chart.

MIRACINAE IN AUSTRALIA: Investigating the biodiversity of an unexplored subfamily of parasitoid wasps

Mollie Slater-Baker, Erinn Fagan-Jeffries
University of Adelaide

MIRACINAE

Miracinae is small subfamily of braconid wasps, characterised by their 14-segmented antennae, a distinctive 'Y' shaped structure on the abdomen, and reduced wing venation. They are parasitoids of the leaf-mining larvae of small moths, and are known to inject their host with symbiotic polydnaviruses when the female lays her eggs, acting to suppress the host immune system.

Actual size!

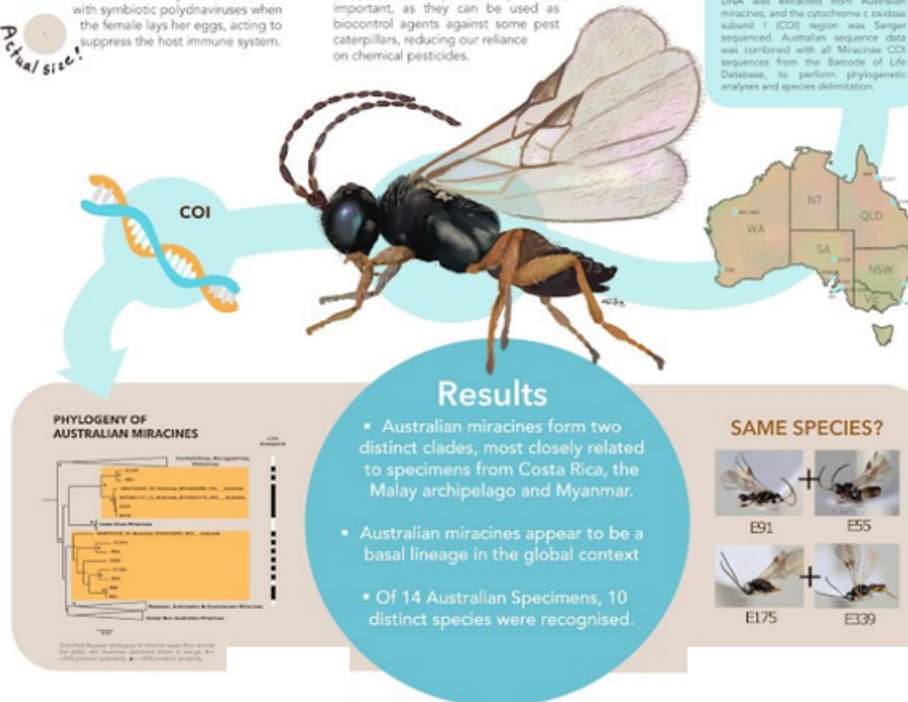
WHY STUDY THEM?

Research on Miracinae is currently very limited, and to date, no work has been done on these wasps in Australia. Describing and studying biodiversity is important for a variety of reasons, from conservation to human health. Research into parasitoid wasps may be particularly important, as they can be used as biocontrol agents against some pest caterpillars, reducing our reliance on chemical pesticides.

AIM: To perform a preliminary exploration of the biodiversity of Australian miracine wasps.

METHODS

DNA was extracted from Australian miracines, and the cytochrome c oxidase subunit I (COI) region was Sanger sequenced. Australian sequence data was combined with all Miracinae COI sequences from the Barcode of Life Database, to perform phylogenetic analyses and species delimitation.



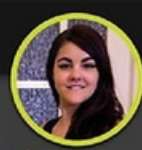
CONCLUSIONS

- The number of distinct specimens in a limited sample suggests that Australia may have many endemic species of Miracinae
- Morphological differences between delimited species were difficult to characterise, and some species may be morphologically cryptic
- DNA extraction was unsuccessful for older samples collected in 2005, so specimen age may be a limiting factor for DNA sequencing

WHAT NEXT?

There is much more to be done towards understanding the Australian Miracinae, and further avenues of study may involve:

- Increased sampling and barcoding of miracines from across Australia
- Phylogenetic analysis using additional loci such as internal transcribed spacer 2 (ITS2) and wingless (WG) to produce a more robust phylogeny for the group
- Detailed imaging of morphological features using scanning electron microscopy
- Formal identification and description of Australian species



SIZING UP THE CRAB INVADERS!

Morphometrics for *Carcinus* spp. can assess population variability & may assist with species identification



@reneccampbell

René Campbell¹, Sabine Dittmann¹, Michael Gardner¹, Marty Deveney²

rene.campbell@flinders.edu.au

CRUSTACEAN DOMINATION

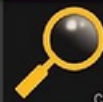
European shore crabs - some of the most widespread marine invasive species worldwide. Two species in the genus:

Carcinus maenas (Atlantic) & *C. aestuarii* (Mediterranean)



Prey on native & commercial species = impacts!

WHICH CRAB IS WHICH?



Taxonomy & population biology: may assist management
Morphometrics: quantitative analysis of shape & form

Carcinus species not confirmed in South Australia

Can assess taxonomic ID & local population variability



Populations: Variation due to life-history or plasticity

Species ID: Potential insight into hybridisation, adaptation or source populations

CRAB COLLECTION

Field sites: Coast of Gulf St Vincent, South Australia

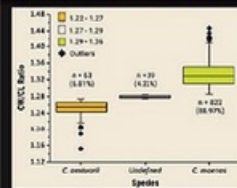
Habitat regions: Rocky shore/ estuary port & mangroves



Specimens: Crabs collected with baited traps & time searches

Seasonal from 2012 - 2018

A LOOK AT SOME DATA



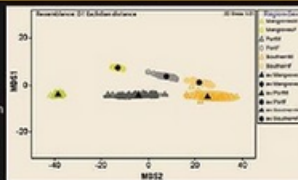
924 *Carcinus* measured (CW/CL ratio)

Higher proportions of *C. maenas* than *C. aestuarii* ($p < 0.05$)

Species not certain as female ratios are more variable

MDS + bootstrap averages for 12 morphometric features

Differences for both sex & region - male crabs most distinct in mangroves



CW & CL - strongest drivers behind these differences

Mangrove crabs tend to be larger = more distinct grouping in MDS

SIMPER Analysis

Mangroves mean similarity	Port mean similarity	Rocky shore mean similarity
85.85%	84.92%	79.67%

Major contributions - carapace width (~16%), carapace length (~14.5%), right chela length (~9.5%)

MEASUREMENTS

Measurements: Digital callipers used for morphometrics

Species ID: Carapace ratios can help identify species (based on previous literature)



Population: Multivariate analyses for morphometrics between regions & sex assessed

WHAT'S NEXT?



Morphometrics - influenced by sex, so species ID inconclusive

Population: Morphology distinct across regions + sex

Potential adaptations to habitats + life history (dispersal)

Population genetics will be assessed - connectivity & gene flow



Species ID: CW/CL ratios suggest both species are present in SA

Sex & intraspecific variability can influence ratios - unreliable?

DNA: species being confirmed with DNA barcoding (COI gene)

IMPLICATIONS

SA's Mediterranean climate - ideal for *C. aestuarii* establishment, or plasticity for *C. maenas*



Habitat use & potential range expansion - concerns for management of this global invader

FLEXIBLE FORAGING BEHAVIOUR IN WILD ZEBRA FINCHES AND ITS RELATION WITH TEMPERATURE

CATERINA FUNGHI, LUKE MCCOWAN, WIEBKE SCHUETT AND SIMON GRIFFITH

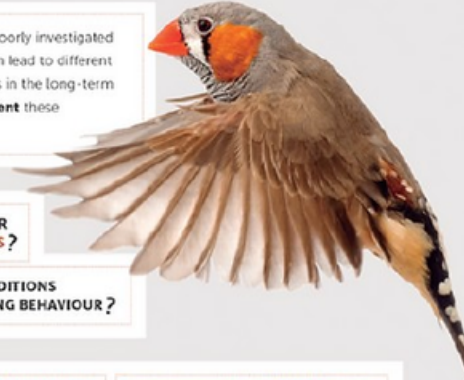
INTRODUCTION

- Foraging as behavioural trait has been poorly investigated
- Fluctuation in environmental conditions can lead to different behavioural strategies being equal in fitness in the long-term
- In extreme and unpredictable environment these fluctuations are more pronounced

QUESTIONS

IS INDIVIDUAL FORAGING BEHAVIOUR
CONSISTENT IN WILD ZEBRA FINCHES?

HOW DO THE ENVIRONMENTAL CONDITIONS
(TEMPERATURE) INFLUENCE FORAGING BEHAVIOUR?



MATERIALS AND METHODS

18 FEEDERS RANDOMLY LOCATED (I.E. 2 TRIALS)

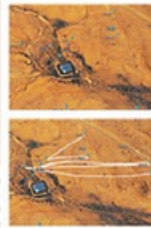


Zebra finches were PIT-tagged and foraging behaviour monitored using a decoder-antenna-feeder system over 3 weeks.

PCA ANALYSIS ON DAILY
FORAGING BEHAVIOUR
OF 72 ADULTS

PCA variation	PCA foraging
Visits	0.3
Distance travelled	0.34
No. feeders	0.5
Feeder fidelity	-0.48

(example of 2 individuals showing different foraging behaviour. Dots represent feeders located 800m around a dam)

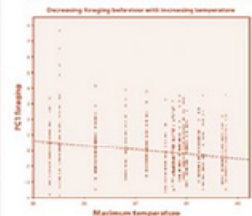


LOW PCA foraging:
• visits per feeder
• distance travelled
• feeders visited
• fidelity

HIGH PCA foraging:
• visits per feeder
• distance travelled
• feeders visited
• fidelity

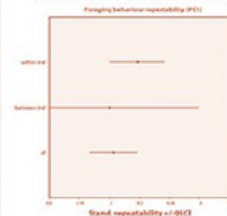
RESULTS AND DISCUSSION

IN EXTREME ENVIRONMENTS FORAGING
BEHAVIOUR WAS INFLUENCED BY TEMPERATURE



(Spearman's $\rho = -0.68$, $N = 68$, Negative relation between foraging behaviour (PCA) and Temperature (°C))

CHANGED ENVIRONMENT LED TO FLEXIBLE
FORAGING BEHAVIOUR - OPPORTUNISM?



(Repeatability (RM), calculated considering all days of all trials, between and within trials)



Contact:
caterina.funghi@students.mq.edu.au

Check Your Poster

Format

One A1-size page, portrait

Group ID and all members' names

Readability

Clear main message in 10 seconds

Readable from 1–2 metres away

Ask a classmate to explain your main message

Revise what they cannot understand

L-6

Unsupervised Learning and Reinforcement Learning I

MAIB7002

Fundamentals of AI Models and Technologies

Three Ways to Learn

What information guides the learning?

Supervised Learning

Examples + **Target Answers**
Predict Target Answers

Unsupervised Learning

Examples **Without Target Answers**
Discover Structure

Reinforcement Learning

Actions + **Reward** Feedback
Learn a Policy

Wholesale Customers

440 Customers · 6 Spending Categories

A Wholesale Distributor

Sells products to hotels, restaurants, cafés and retailers



440 Business Customers

One row records one customer's annual purchases

What Is Measured?

Annual spending across six product categories

No Labels

Spending amounts, not quantities · m.u. = monetary units · UCI Wholesale Customers

Explore the Customer Dataset

Annual Spending (m.u.) · Select a Row · Click a Heading to Sort

Customer	Fresh ↑	Milk ↑	Grocery ↑	Frozen ↑	Paper / Cleaning ↑	Delicatessen ↑
1	12,669	9,656	7,561	214	2,674	1,338
2	7,057	9,810	9,568	1,762	3,293	1,776
3	6,353	8,808	7,684	2,405	3,516	7,844
4	13,265	1,196	4,221	6,404	507	1,788
5	22,615	5,410	7,198	3,915	1,777	5,185
6	9,413	8,259	5,126	666	1,795	1,451

Rows 1–6 of 440

◀ Previous

Next ▶

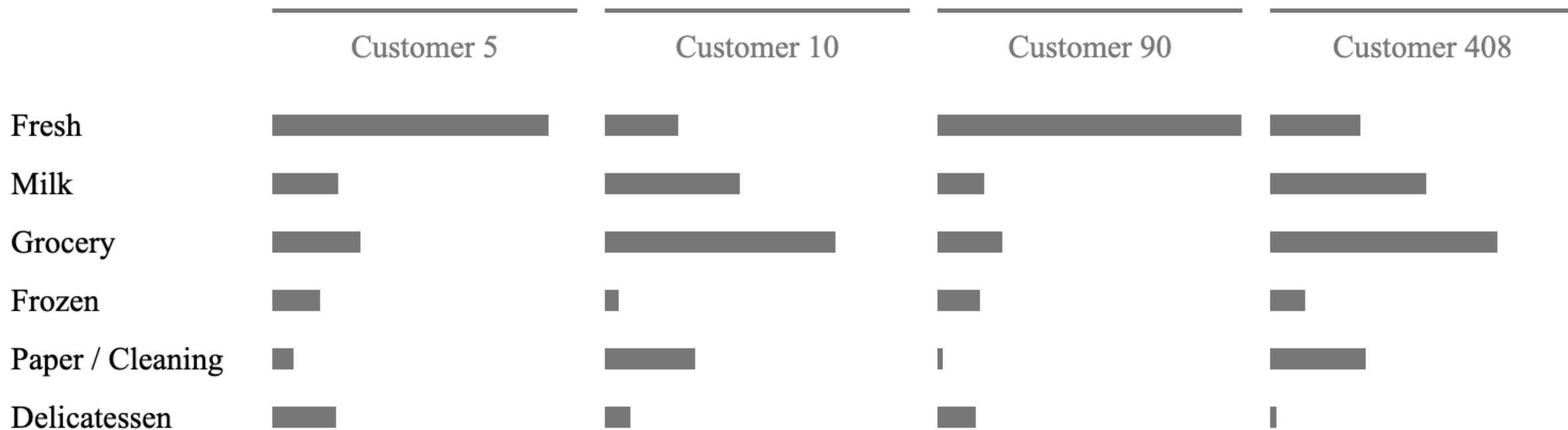
Reset

View Customer Profile

440 wholesale customers · m.u. = monetary units · UCI / Cardoso (2013) · CC BY 4.0

Clustering Customers

Could we organize 440 customers by what they buy?

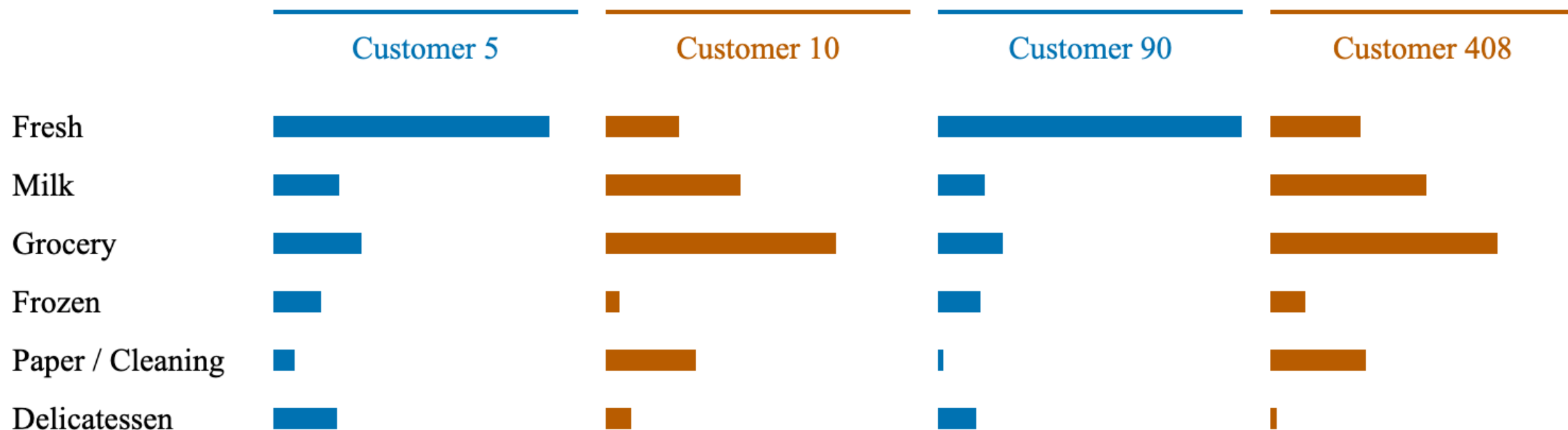


Which Customers Would You Put Together?

4 selected real customers · All bars use the same 0–25,000 m.u. scale · Annual spending

Clustering Customers

Could we organize 440 customers by what they buy?

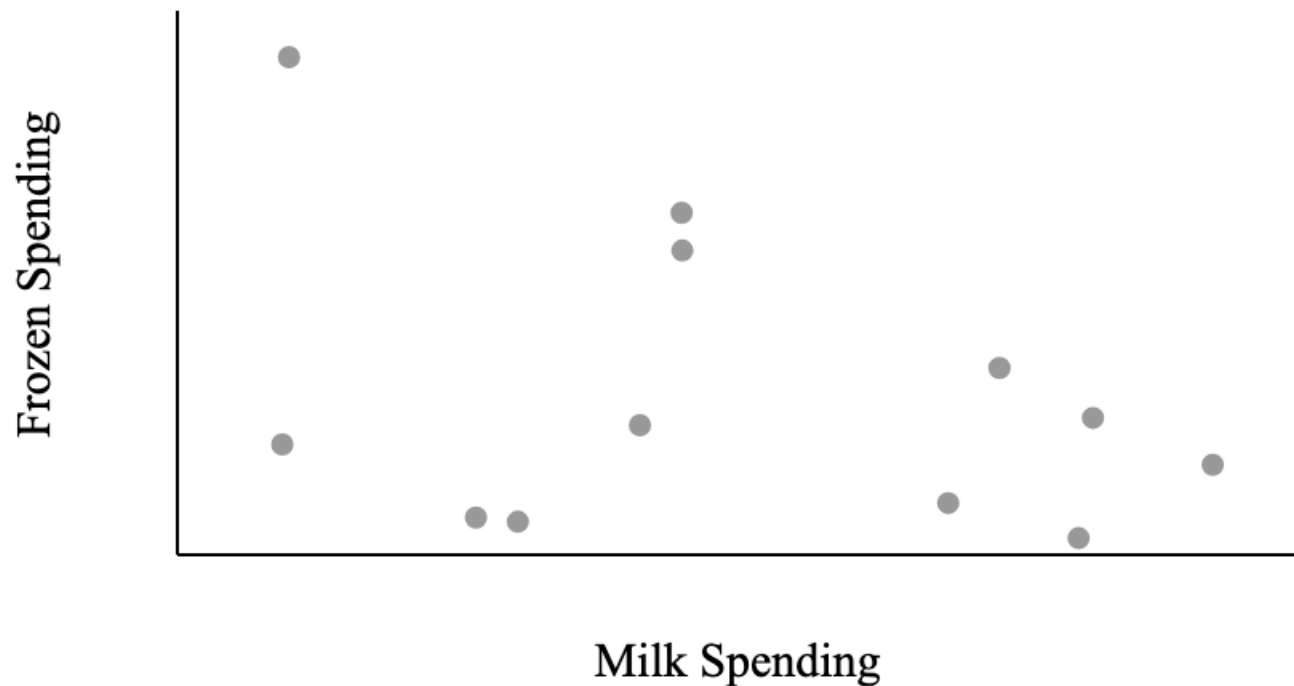


Similar Within Each Group · Different Across Groups

4 selected real customers · All bars use the same 0–25,000 m.u. scale · Annual spending

Visualizing Customer Similarity

Each Dot Represents One Customer



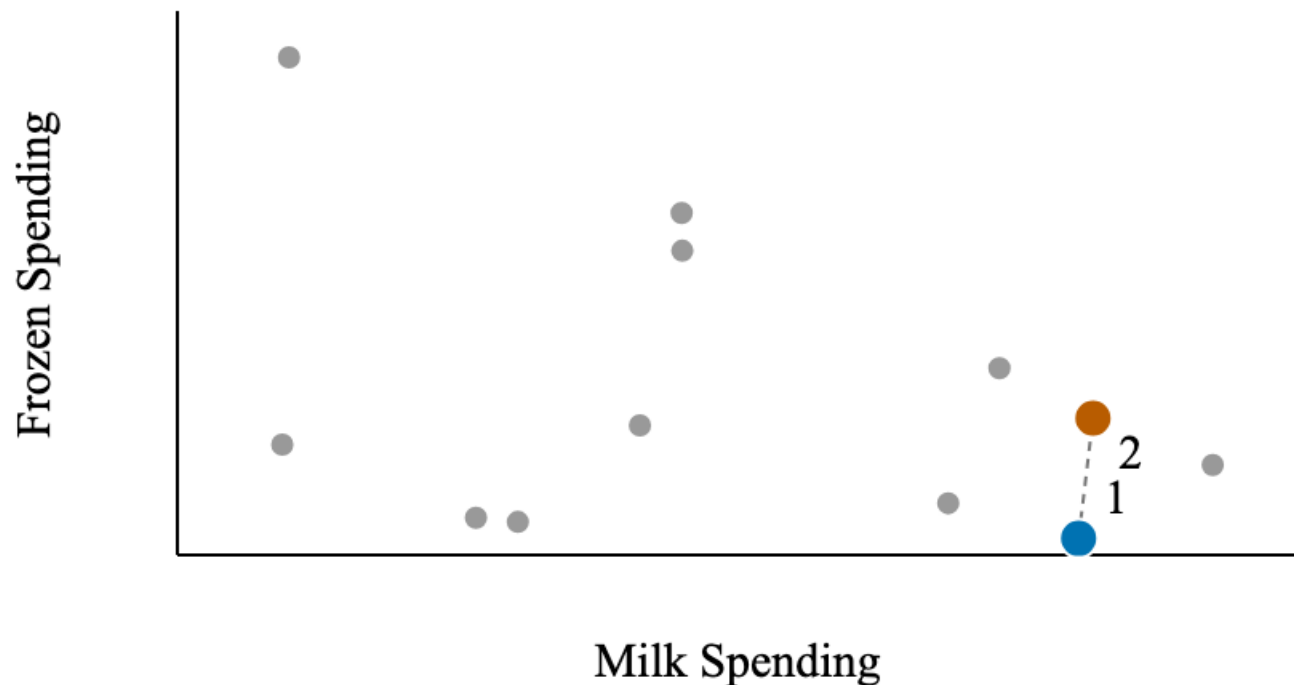
2D: Two Features

Which Customers Look Similar?

First 12 real customers · Annual spending (m.u.)

Visualizing Customer Similarity

Each Dot Represents One Customer



2D: Two Features

Nearby points have similar spending on the features shown

Customers 1 & 2

Customers 5 & 10

First 12 real customers · Annual spending (m.u.)

Visualizing Customer Similarity

Each Dot Represents One Customer



3D: Three Features

Add Grocery Spending

Customers 1 & 2

Customers 5 & 10

Rotate View



First 12 real customers · Annual spending (m.u.)

Visualizing Customer Similarity

Each Dot Represents One Customer



6D: Six Features

Compare all six spending categories

Customers 1 & 2

Customers 5 & 10

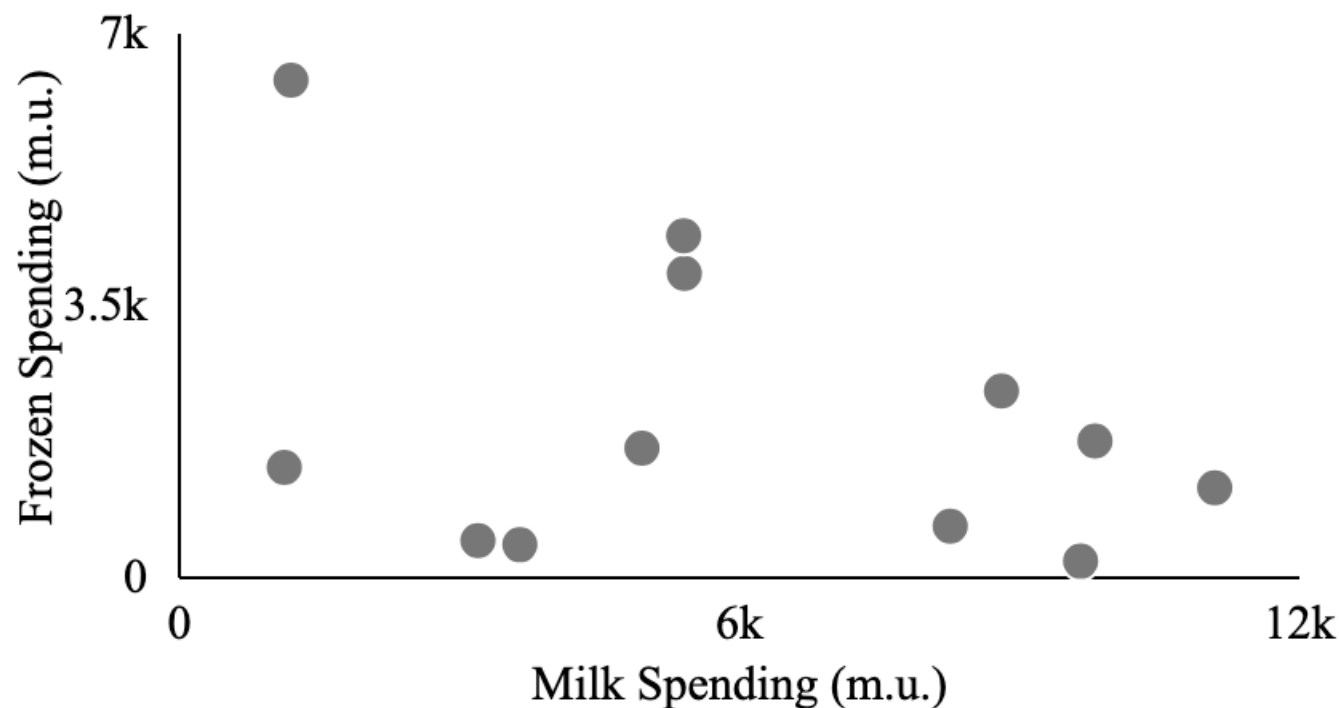
Rotate View



First 12 real customers · Annual spending (m.u.)

Applications of Clustering

What could we group, and why would it help?



Customer Purchases

Which customers buy similarly?

Customers

Complaints

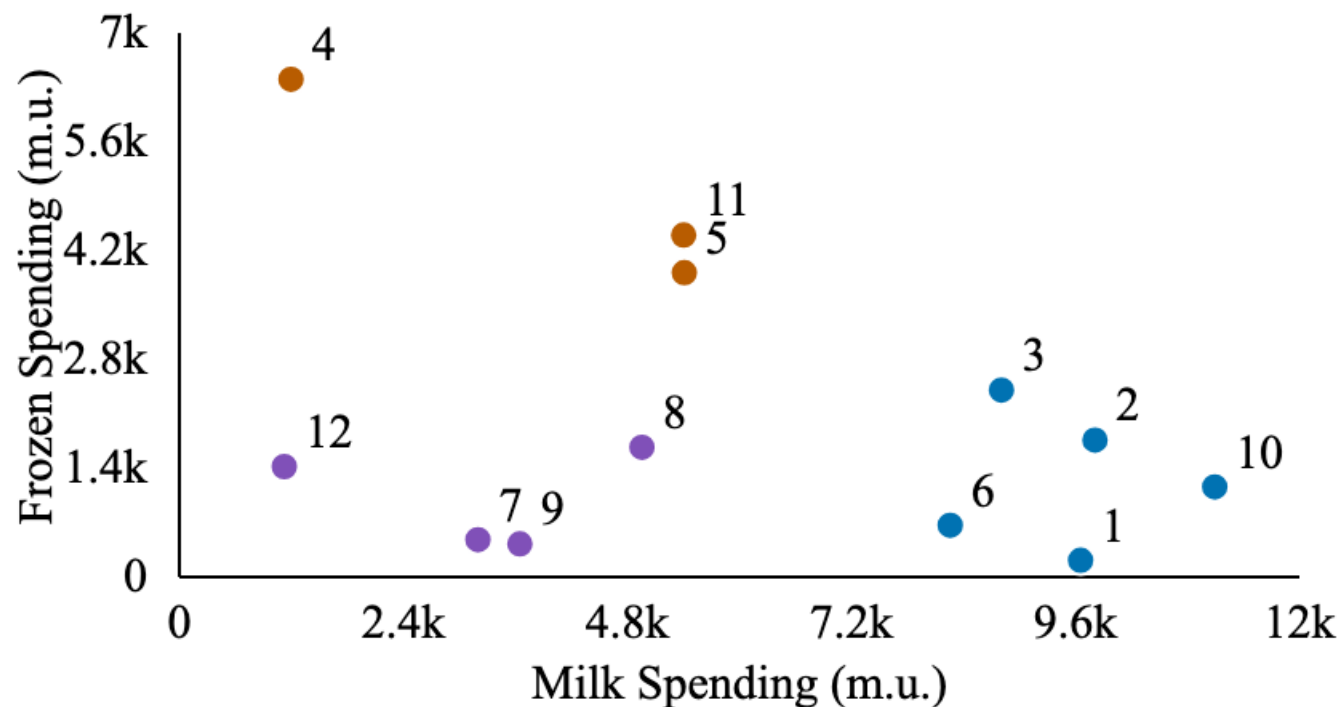
Stores

Reveal Groups

First 12 real customers · Milk / Frozen · Grouping uses scaled features

Evaluating Clusters

Which Grouping Makes More Sense?



Compare the Four Groupings

Which would you trust, and why?

Grouping A

Grouping B

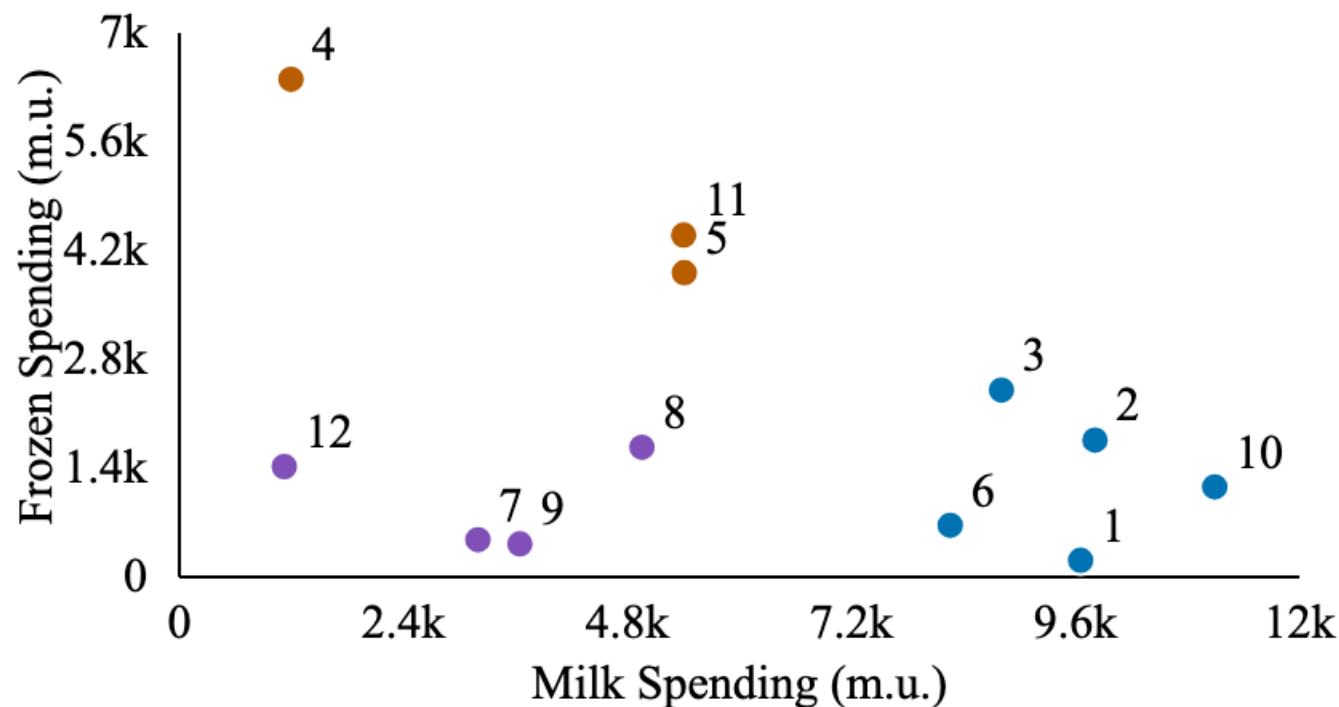
Grouping C

Grouping D

Same 12 customers · Three groups of the same sizes · Only membership changes

Evaluating Clusters

Which Grouping Makes More Sense?



Check Separation

Similar within groups, different between groups

Test Business Usefulness

Grouping A

Grouping B

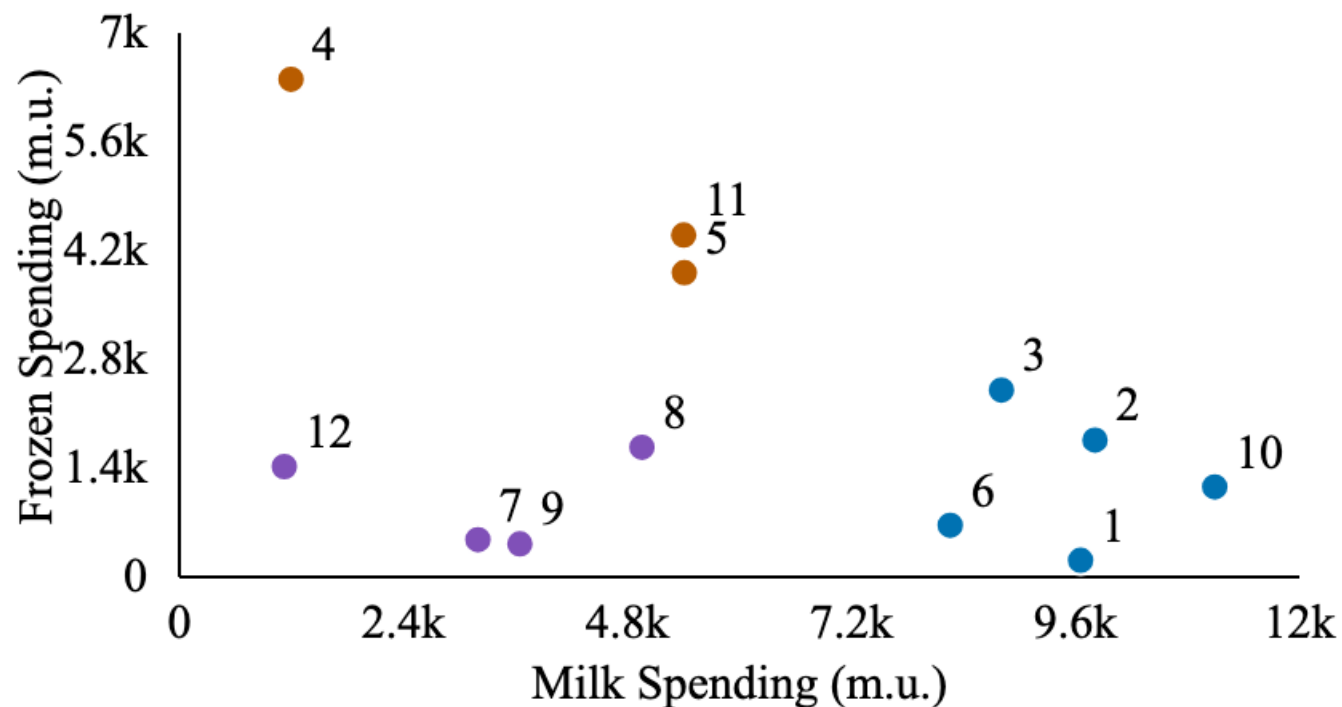
Grouping C

Grouping D

Same 12 customers · Three groups of the same sizes · Only membership changes

Number of Groups

How Many Groups Would Be Useful?



More Groups Give Finer Detail

When would that extra detail be useful?

2 Groups

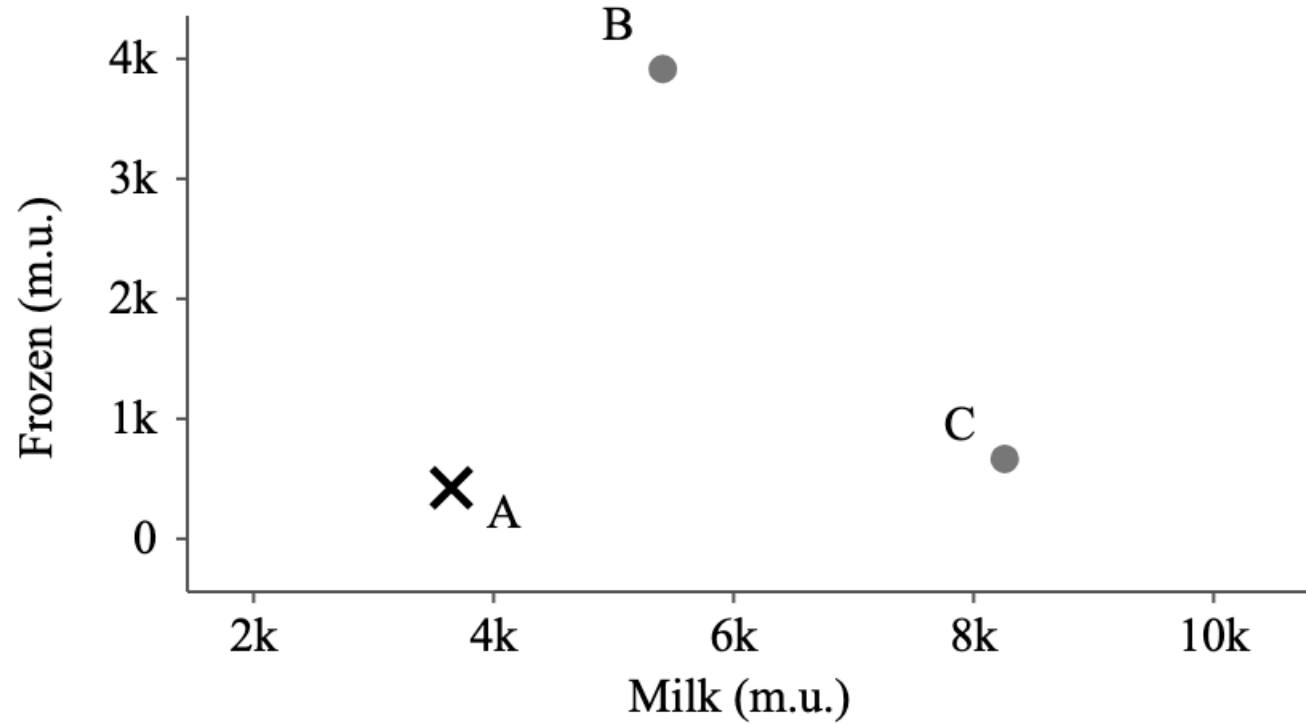
3 Groups

4 Groups

Same 12 customers · Milk / Frozen · More groups are not automatically better

Similarity and Feature Scaling

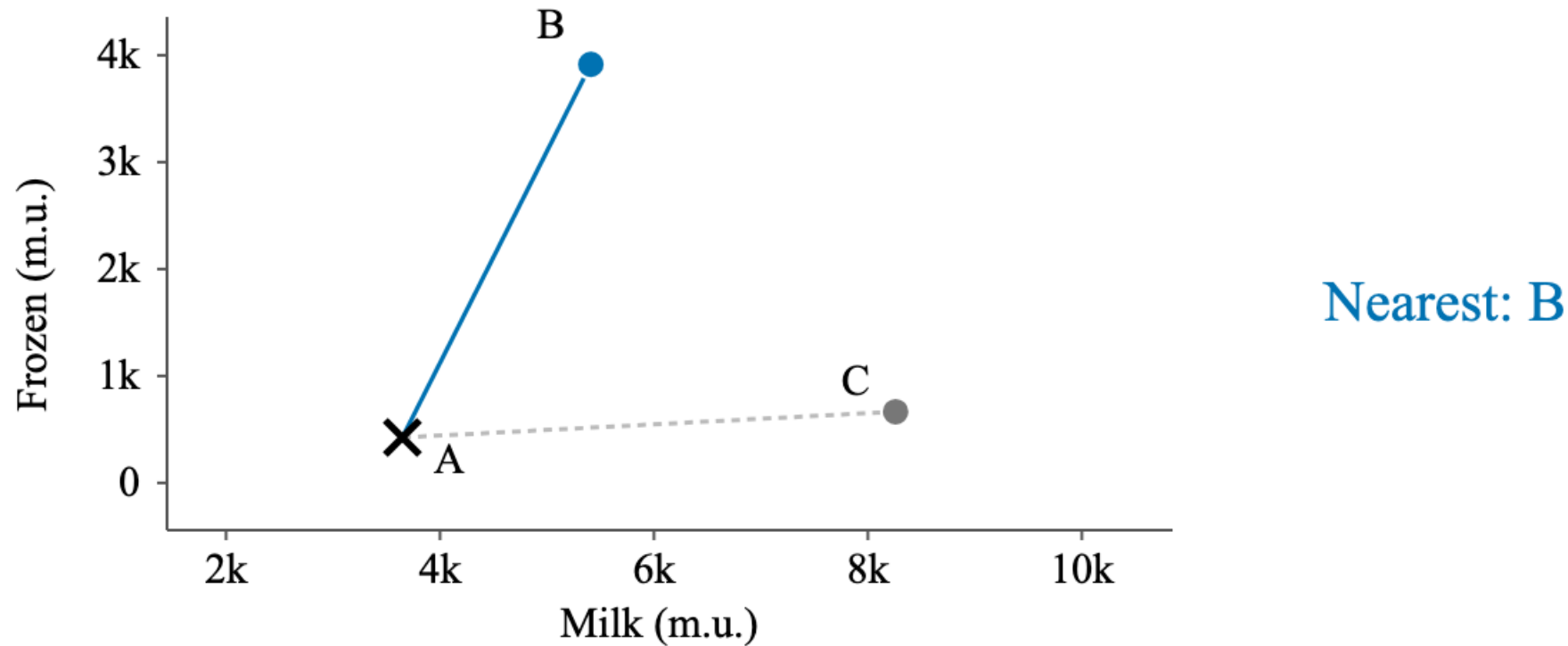
Which Customer Is Most Similar to A?



Scaling: All 440 Customers · Three Shown: A = 9, B = 5, C = 6

Similarity and Feature Scaling

Which Customer Is Most Similar to A?



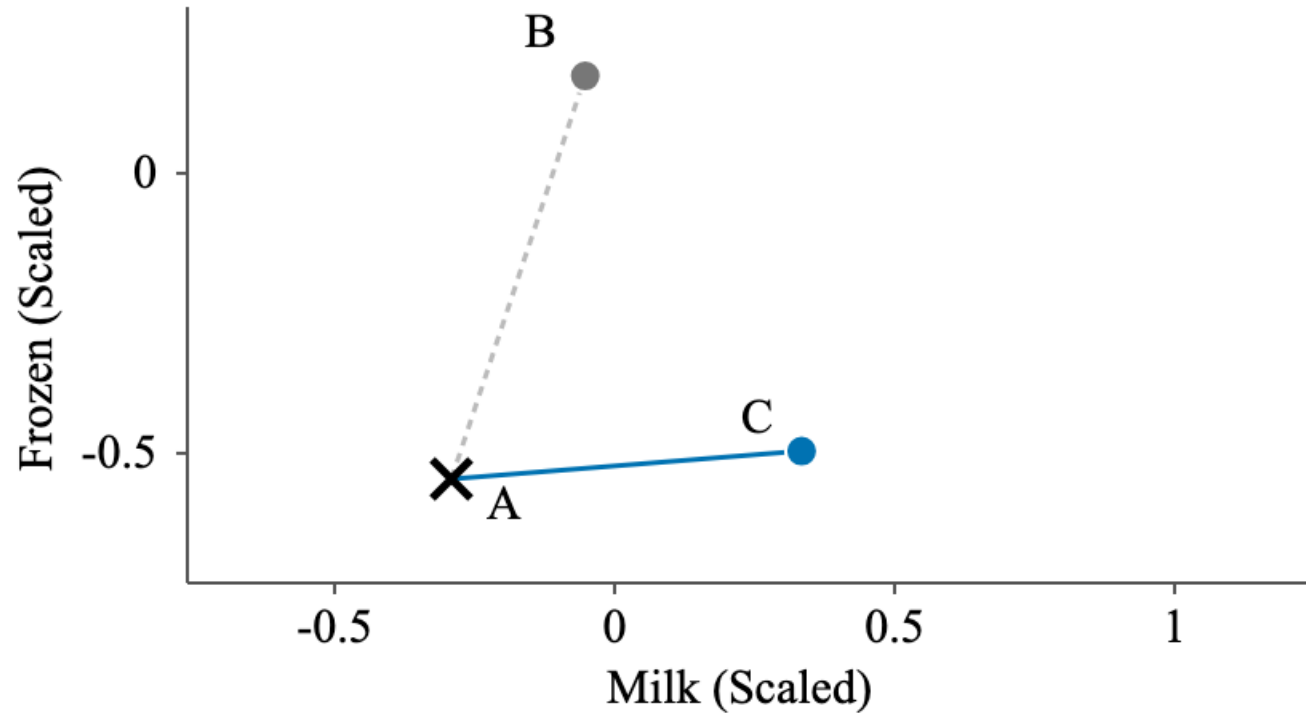
Original Units

Comparable Scales

Scaling: All 440 Customers · Three Shown: A = 9, B = 5, C = 6

Similarity and Feature Scaling

Which Customer Is Most Similar to A?



Standardize Each Feature

Subtract Its Mean

Divide by Its Standard Deviation

Mean = 0

Standard Deviation = 1

Nearest: C

Original Units

Comparable Scales

Scaling: All 440 Customers · Three Shown: A = 9, B = 5, C = 6

Clustering Methods

Different Ways to Form Groups

K-Means



Groups Around Centres

Hierarchical



Nested Groups

DBSCAN



Dense Regions

Illustrative diagrams · Different methods use different ideas of a group

Clustering Methods

Different Ways to Form Groups

K-Means



Groups Around Centres

Hierarchical



Nested Groups

DBSCAN



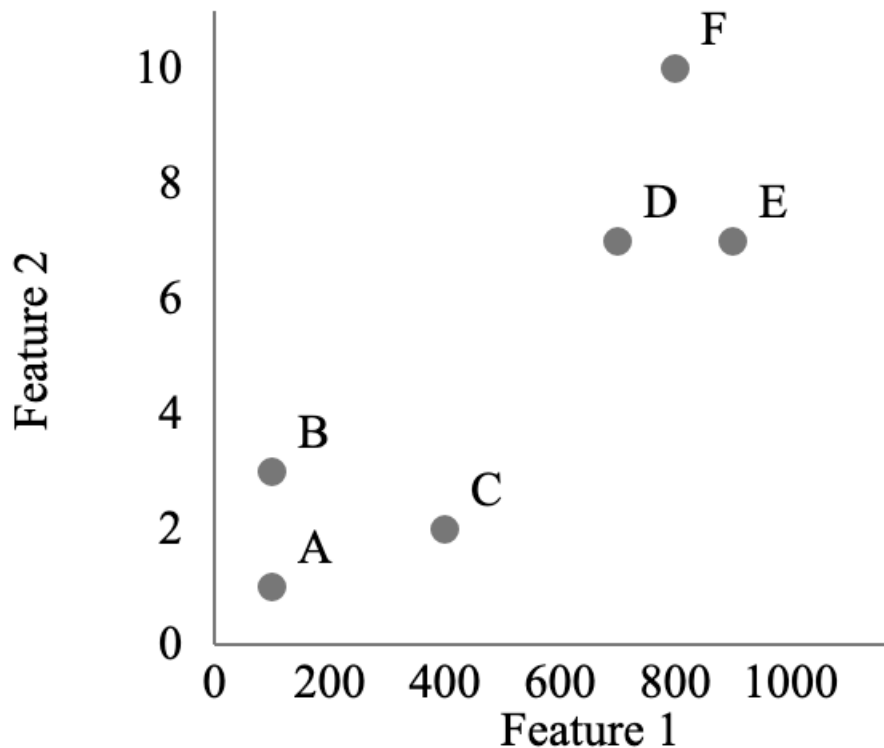
Dense Regions

We'll Explore K-Means

Illustrative diagrams · Different methods use different ideas of a group

K-Means Clustering

How Would You Group These Points?



1. Scale the Features

2. Choose k

3. Pick Random Starting Centres

4. Assign to the Nearest Centre

5. Calculate Each Group's Mean

Repeat Steps 4 and 5 Until Nothing Changes

Scale Features

2

3

4

Pick Centres

Assign Points

Calculate New Centres

Set 1

Set 2

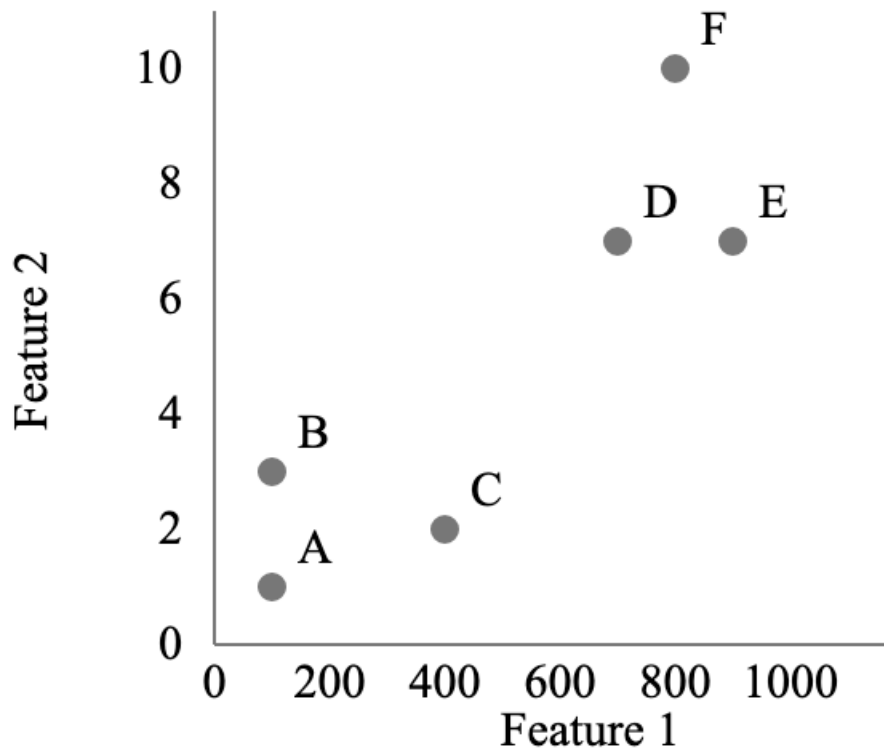
Set 3

Set 4

Reset Example

Choosing Starting Centres

How Would You Group These Points?



1. Scale the Features

2. Choose k

3. Click to Place 2 Centres

4. Assign to the Nearest Centre

5. Calculate Each Group's Mean

Repeat Steps 4 and 5 Until Nothing Changes

Scale Features

2

3

4

Choose Again

Assign Points

Calculate New Centres

Set 1

Set 2

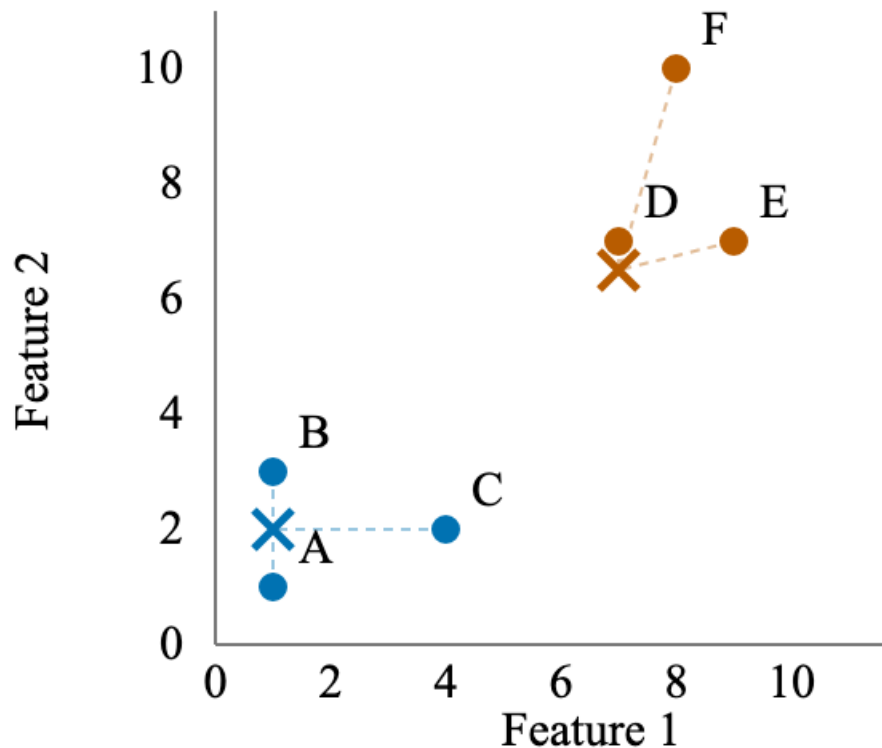
Set 3

Set 4

Reset Example

Calculating New Centres

What Is the Mean of the Three Blue Points?



Look at A, B and C

A: (1, 1)

B: (1, 3)

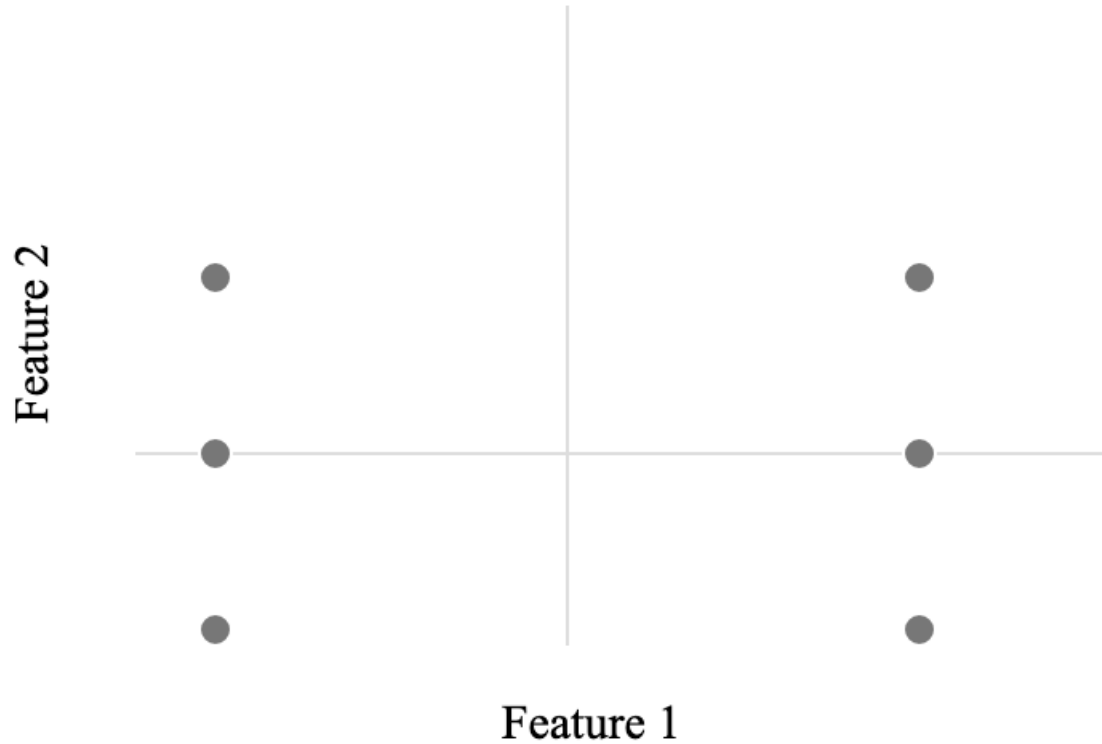
C: (4, 2)

What Are the Mean x and y?

Show New Centres

Task: Starting Centres

Can Different Starting Centres Lead to Different Final Groups?



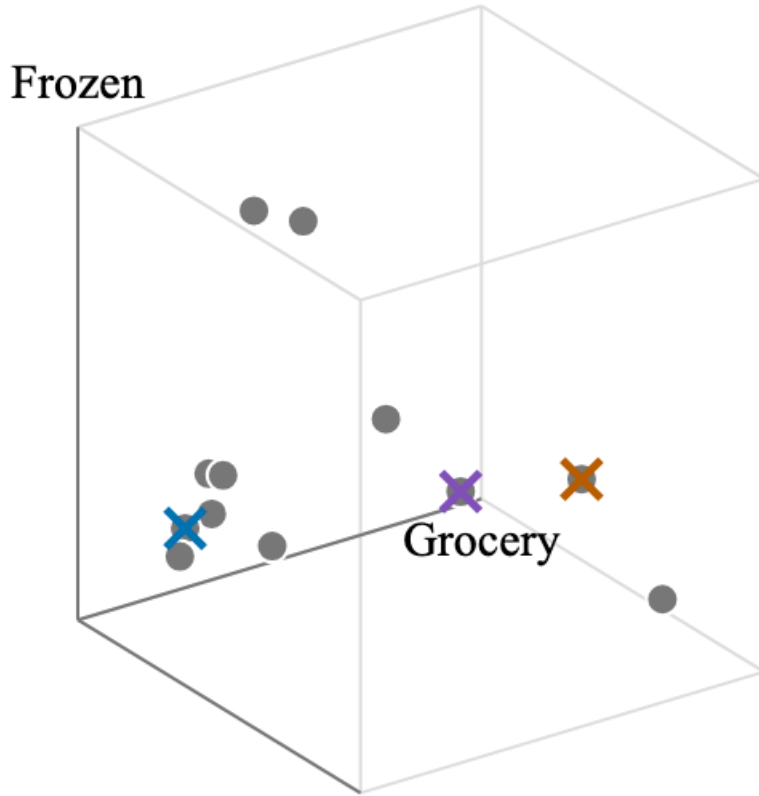
Same Data · Same $k = 2$

Submit
Yes or No

Explore After Voting

Customer Groups in 3D

12 Selected Customers · Three Spending Features · $k = 3$



Same Algorithm · Three Features

Find Each Customer's Nearest Centre

Assign Points

Distances Use All Three Features

Rotate the View to Inspect the Groups

View Spending Profiles

Paper / Cleaning

Rotate View



Restart

K-Means Notebook Task

Which Measurements Help Us Find Groups?

Predict, Then Try

Explain Your Result

Petals or Sepals?

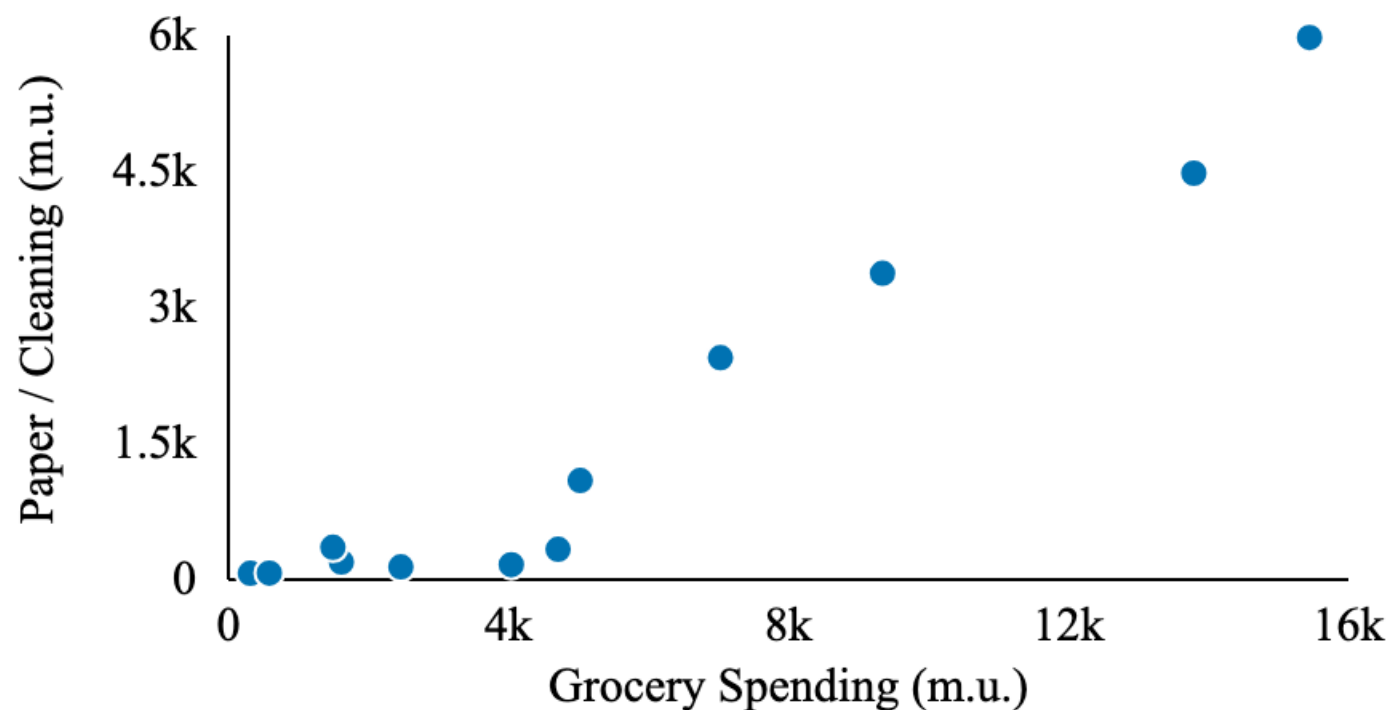
Which Groups Look More Separated?

Keep $k = 3$ · Change the Feature Pair

maib7002.dirk.hk/L6.zip

Six Measurements, One View

12 Customers · Each Dot Is One Customer



Six Features per Customer

Fresh

Milk

Grocery

Frozen

Paper / Cleaning

Delicatessen

Can We Summarize 6D in 2D?

Dimensionality Reduction

6D → 2D

Grocery + Paper / Cleaning

Grocery + Frozen

Milk + Grocery

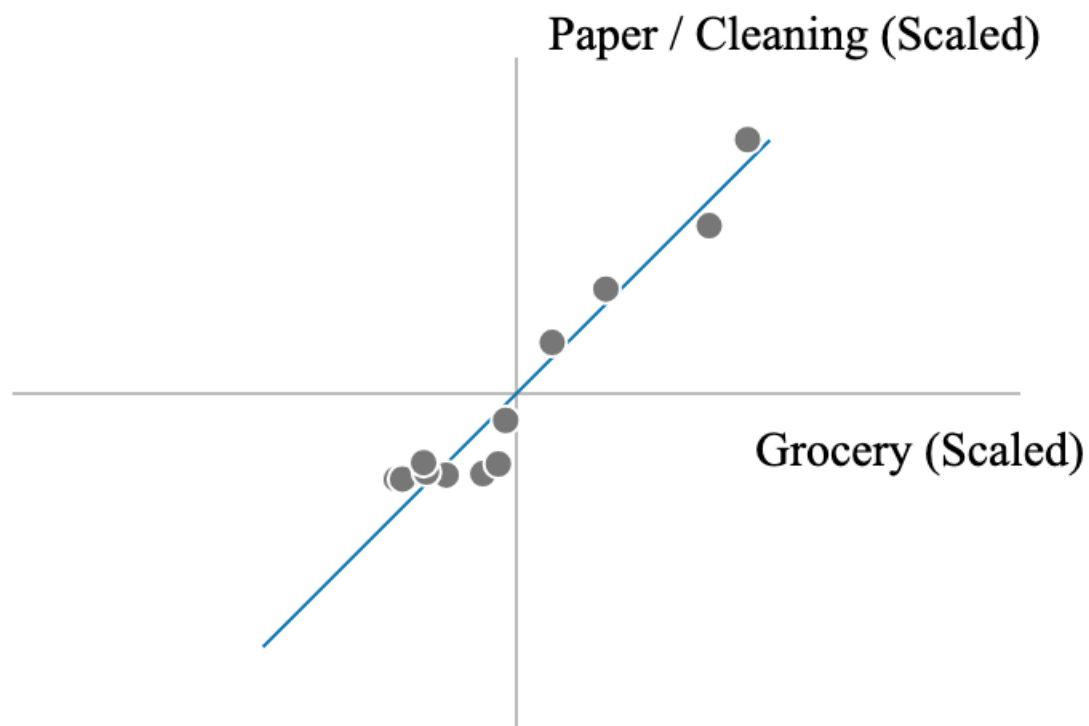
Fresh + Frozen

Milk + Delicatessen

Fresh + Grocery

Related Measurements

A Simpler Example: 2D $\rightarrow$ 1D



Grey: Original Customers

Can One Number Summarize
These Two Measurements?

Rotate Line



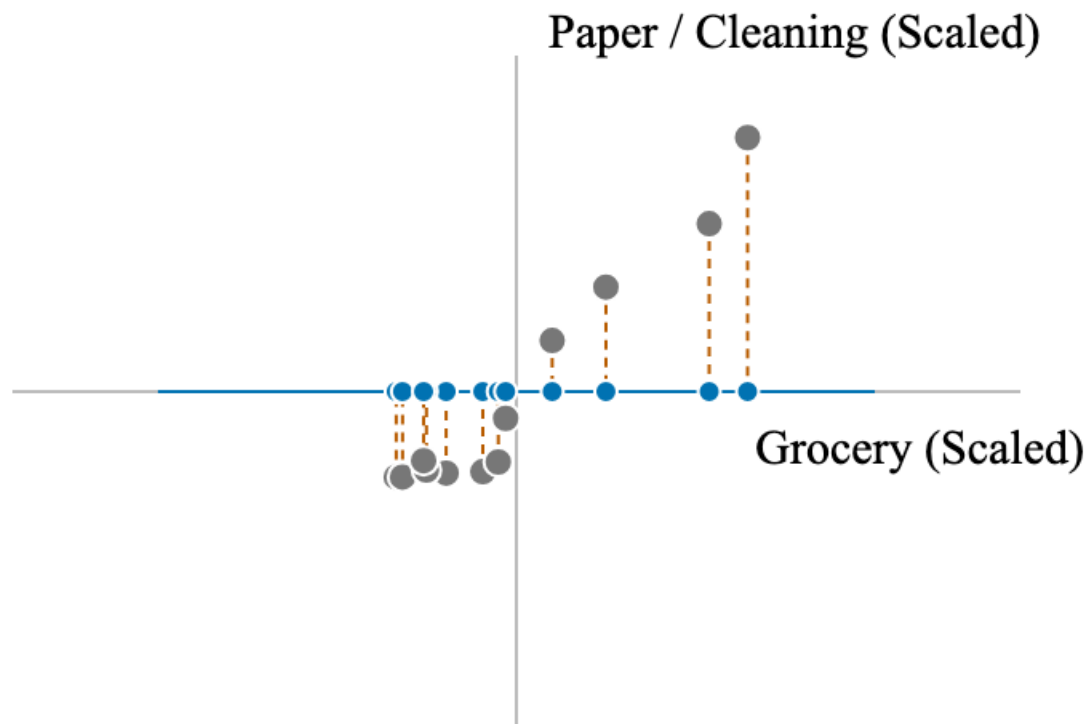
Show Projections

Show 1D

Reset

Principal Component Analysis

How Do We Choose the Best Direction?



Which Direction Keeps
the Most Spread?

Spread Preserved: 50.0%



Variance Measures How
Spread Out the Points Are

Rotate Line



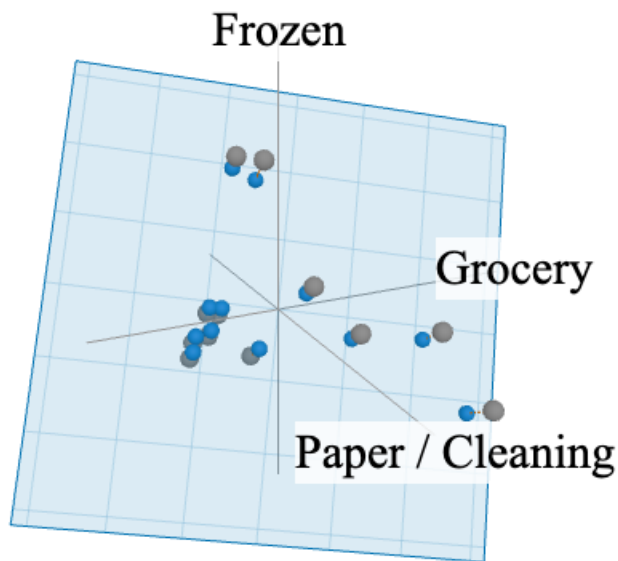
Find Best Direction

Reset

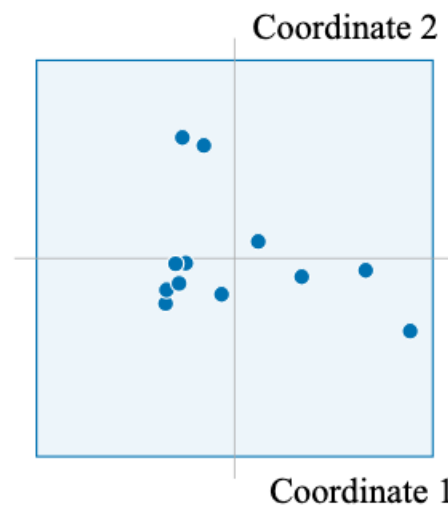
From Three Measurements to Two

3D $\rightarrow$ 2D · 12 Customers · Scaled Features

Customers and Plane in 3D



View on the Plane



Spread Preserved: 82.4%

Two Coordinates per Customer

Drag to Rotate the View · Scroll to Zoom

Tilt Plane



Turn Plane



Hide Projections

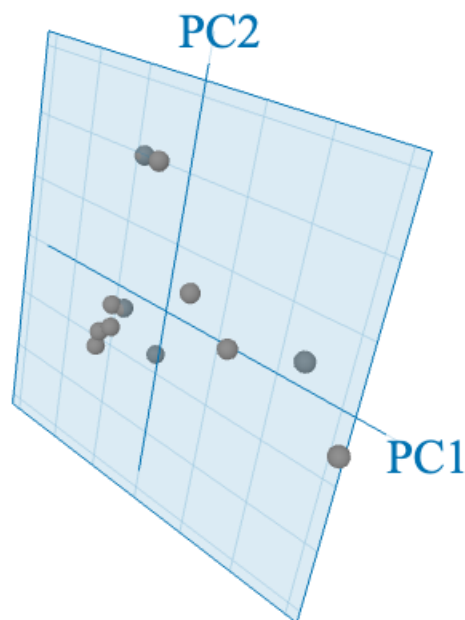
Find Best Plane

Reset

Which Dimensions Do We Keep?

Keep the Directions with the Most Variation

New Coordinate Directions



Variation in Each Component



Components Are New Coordinates

Drag to Rotate the View · Scroll to Zoom

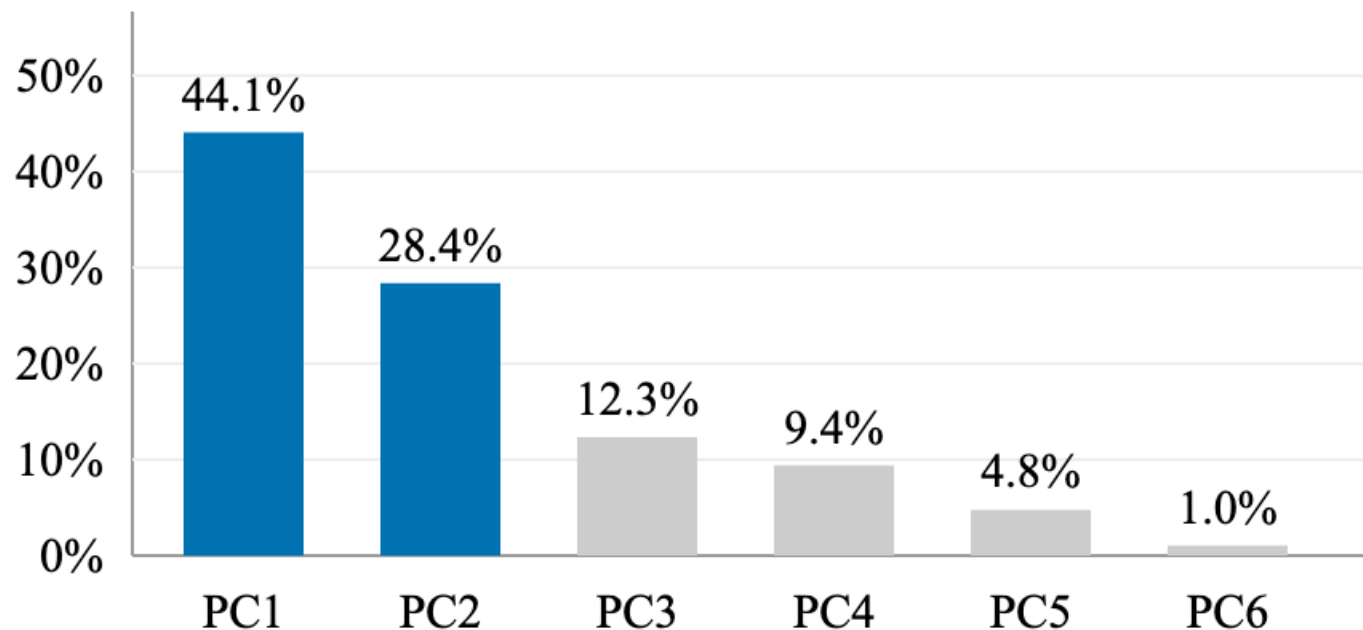
Reveal PC3

Keep Two Components

Reset

How Many Components Should We Keep?

440 Customers · Six Scaled Spending Features



Variation Preserved

72.5%

2 of 6 Components Kept

Is Two Enough?

Two Coordinates Make a 2D Plot
More Components Preserve More Detail

Components to Keep

1

2

3

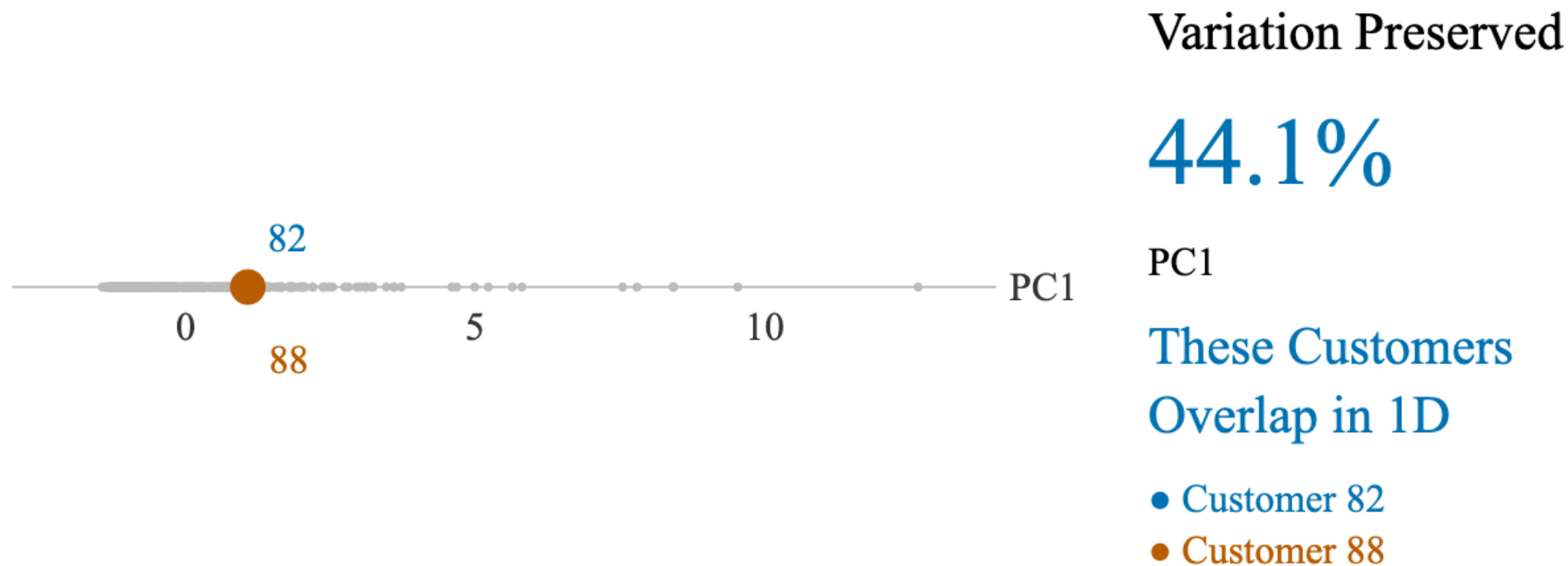
4

5

6

What Does Another Component Add?

440 Customers · The Same PCA, More Coordinates



1D: PC1

2D: Add PC2

3D: Add PC3

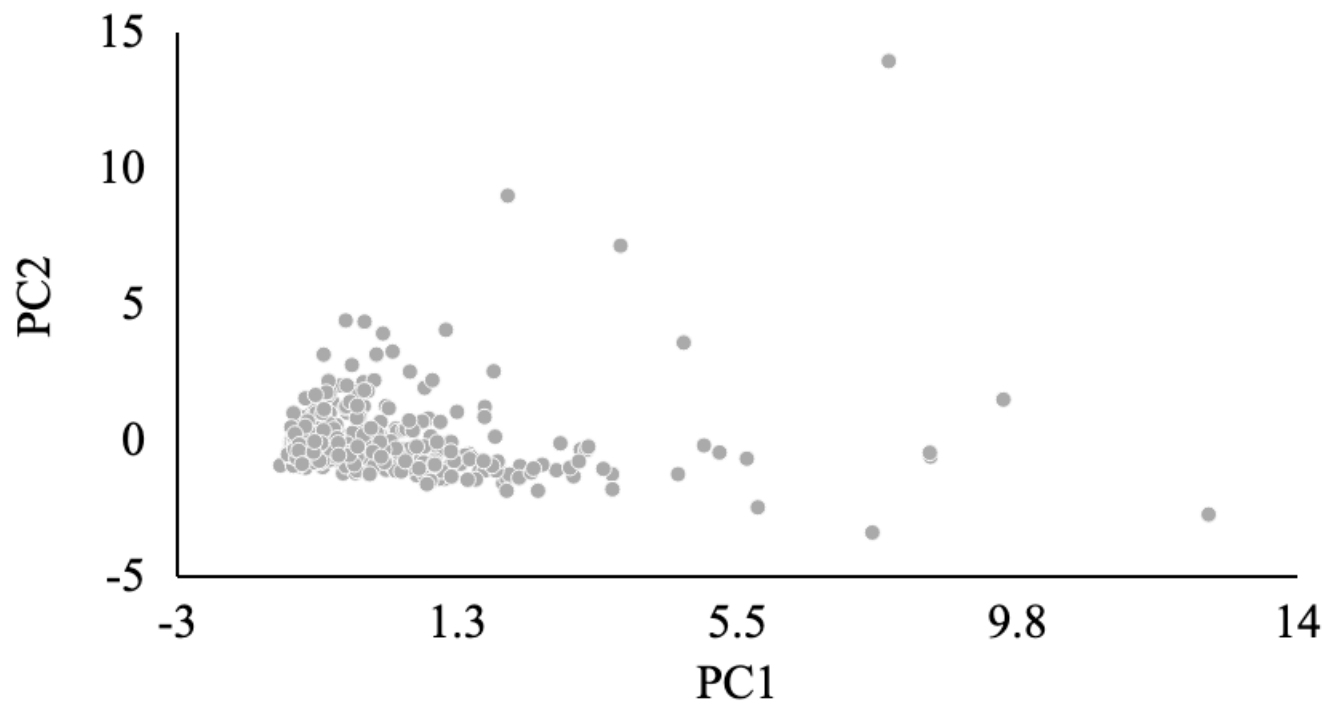
Reset

Seeing the Customer Groups

Six Scaled Features · PCA Positions · K-Means Groups

440 of 440 Customers Visible

Position: PCA



Can We See the
Customer Groups?

Click a Customer to Explore
Their Spending

Show K-Means Groups

Zoom to Centre

Reset

PCA Notebook Task

When Is One Coordinate Enough?

Try One Coordinate

Grocery + Paper/Cleaning

Replace Paper/Cleaning with Frozen

Explain Your Result

Which Pair Loses Less Detail?

maib7002.dirk.hk/L6.zip

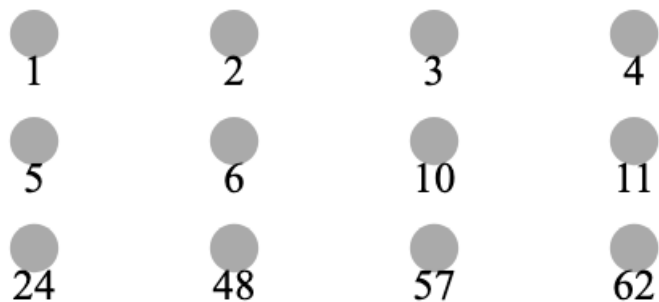
Clustering and PCA

Group Customers with
Similar Buying Patterns

K-Means

PCA

Choose a Method



Customer IDs · $k = 3$

Summarize Six Spending Features
in Two Coordinates

K-Means

PCA

Choose a Method

6D → 2D

Reset

L-6

Unsupervised Learning and **Reinforcement Learning I**

MAIB7002
Fundamentals of AI Models and Technologies

Three Ways to Learn

Now: Learning from Actions and Rewards

Supervised Learning

Examples + **Target Answers**
Predict Target Answers

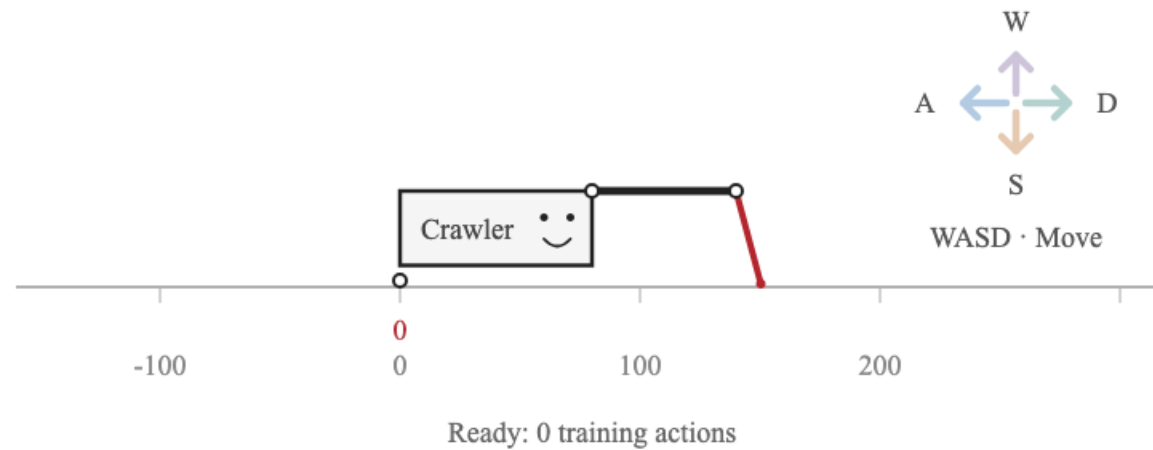
Unsupervised Learning

Examples **Without Target Answers**
Discover Structure

Reinforcement Learning

Actions + **Reward** Feedback
Learn Which Actions to Take

Learning Through Rewards



The Two-Armed Bandit

Which Machine Would You Choose?



Play Blue

Plays: 0 · Total: \$0

Average per Play: —

What Can We Learn by Playing?



Play Red

Plays: 0 · Total: \$0

Average per Play: —

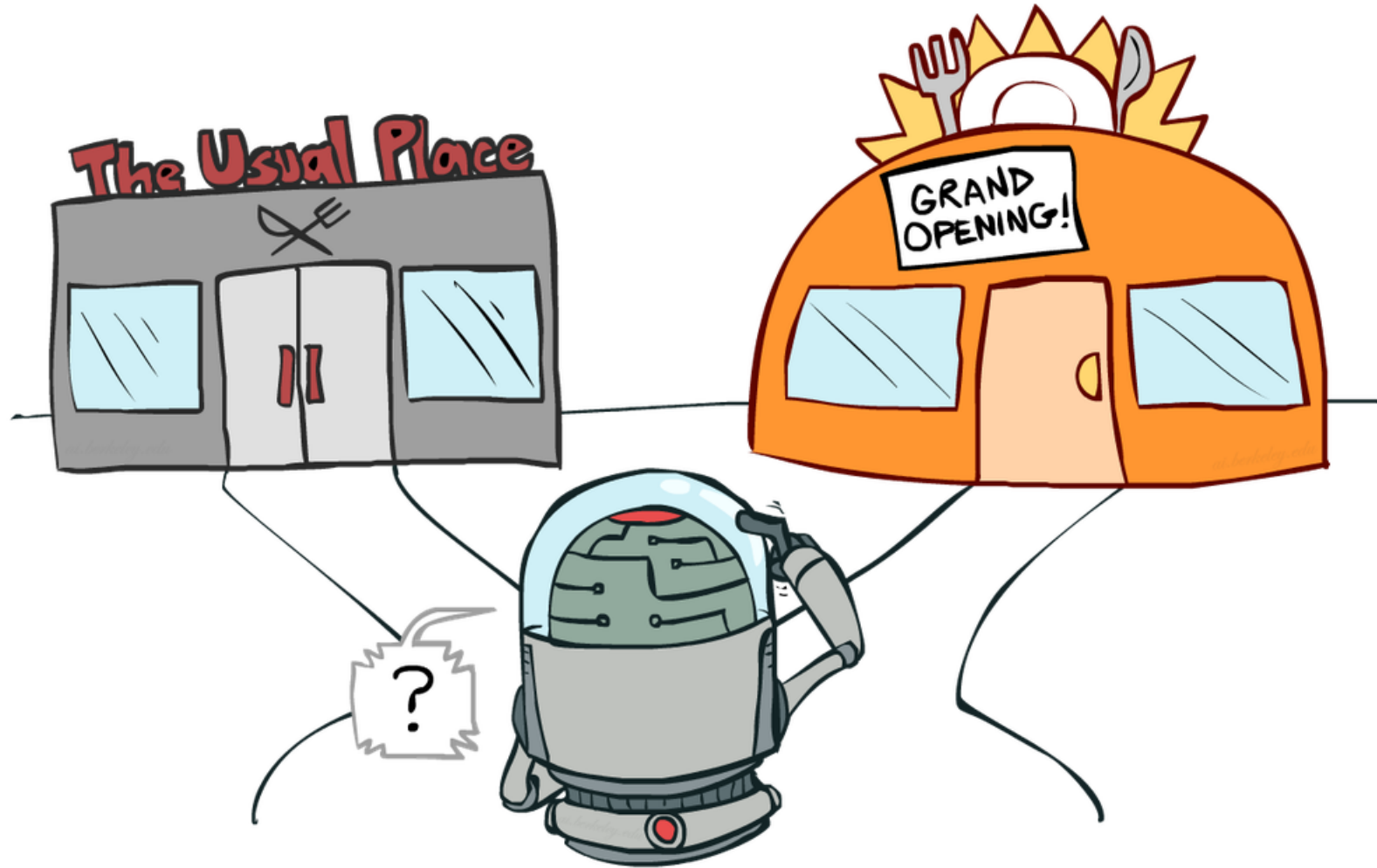
What Can We Learn by Playing?

Choose a Machine

Reveal Chances

New Round

Exploration and Exploitation



Exploration and Exploitation in Practice

What Would You Explore, and What Would You Exploit?

Explore

Exploit

Restaurants

?

?

Online Advertising

?

?

Oil Drilling

?

?

Learn About Alternatives or Use What Works Best So Far?

Reveal All

Learning Which Advert to Show

The Learner Chooses Which Advert to Show

Customer Visits: 0

Grocery Advert

Frozen-Food Advert

0 Purchases / 0 Times Shown

Observed Purchase Rate: —

0 Purchases / 0 Times Shown

Observed Purchase Rate: —

Next Customer → Learner Chooses an Advert

Purchase = Reward 1 · No Purchase = Reward 0

Learn Which Advert Earns More Reward

Keep Exploring While Using What We Have Learned

Next Customer

Reset

Simulated Responses

Exit Ticket

Exit Ticket

No Exit Ticket 😊