Inf2b Learning and Data

http://www.inf.ed.ac.uk/teaching/courses/inf2b/

Lecture 3

Clustering, Collaborative counting review

Iain Murray, 2013

School of Informatics, University of Edinburgh

How to stay on the road?

Self-driving car in the desert:

— You can't trust GPS+map
— Laser range finders can get confused or go off-line
— You have a camera, but. . .
— Off-road in place A looks like on-road in place B

`http://robots.stanford.edu/talks/stanley/`

Today's Schedule:

- — Collaborative counting (review)
- — Clustering
- — How to stay on the road (time allowing)

Review: the confection

m&m's
(185g)

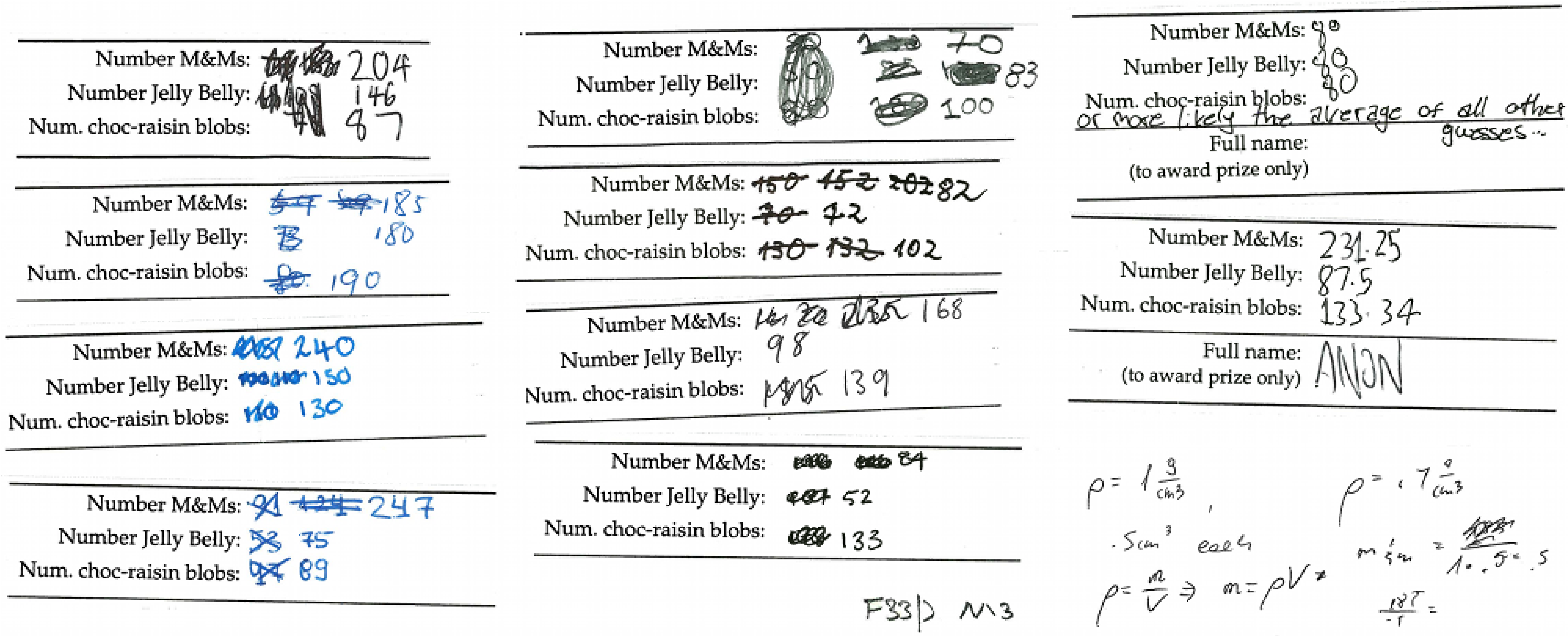
Jelly Belly
(100g)

Chocolate Raisins
(200g)

The importance of guessing

http://StreetFightingMath.com/

Stuff Inf2b students wrote



Number M&Ms: 204
Number Jelly Belly: 146
Num. choc-raisin blobs: 87

Number M&Ms: 185
Number Jelly Belly: 180
Num. choc-raisin blobs: 190

Number M&Ms: 240
Number Jelly Belly: 150
Num. choc-raisin blobs: 130

Number M&Ms: 247
Number Jelly Belly: 75
Num. choc-raisin blobs: 89

Number M&Ms: 70
Number Jelly Belly: 83
Num. choc-raisin blobs: 100

Number M&Ms: 82
Number Jelly Belly: 42
Num. choc-raisin blobs: 102

Number M&Ms: 168
Number Jelly Belly: 98
Num. choc-raisin blobs: 139

Number M&Ms: 84
Number Jelly Belly: 52
Num. choc-raisin blobs: 133

F33D M13

Number M&Ms: 90
Number Jelly Belly: 80
Num. choc-raisin blobs: 80
or more likely the average of all other guesses...
Full name:
(to award prize only)

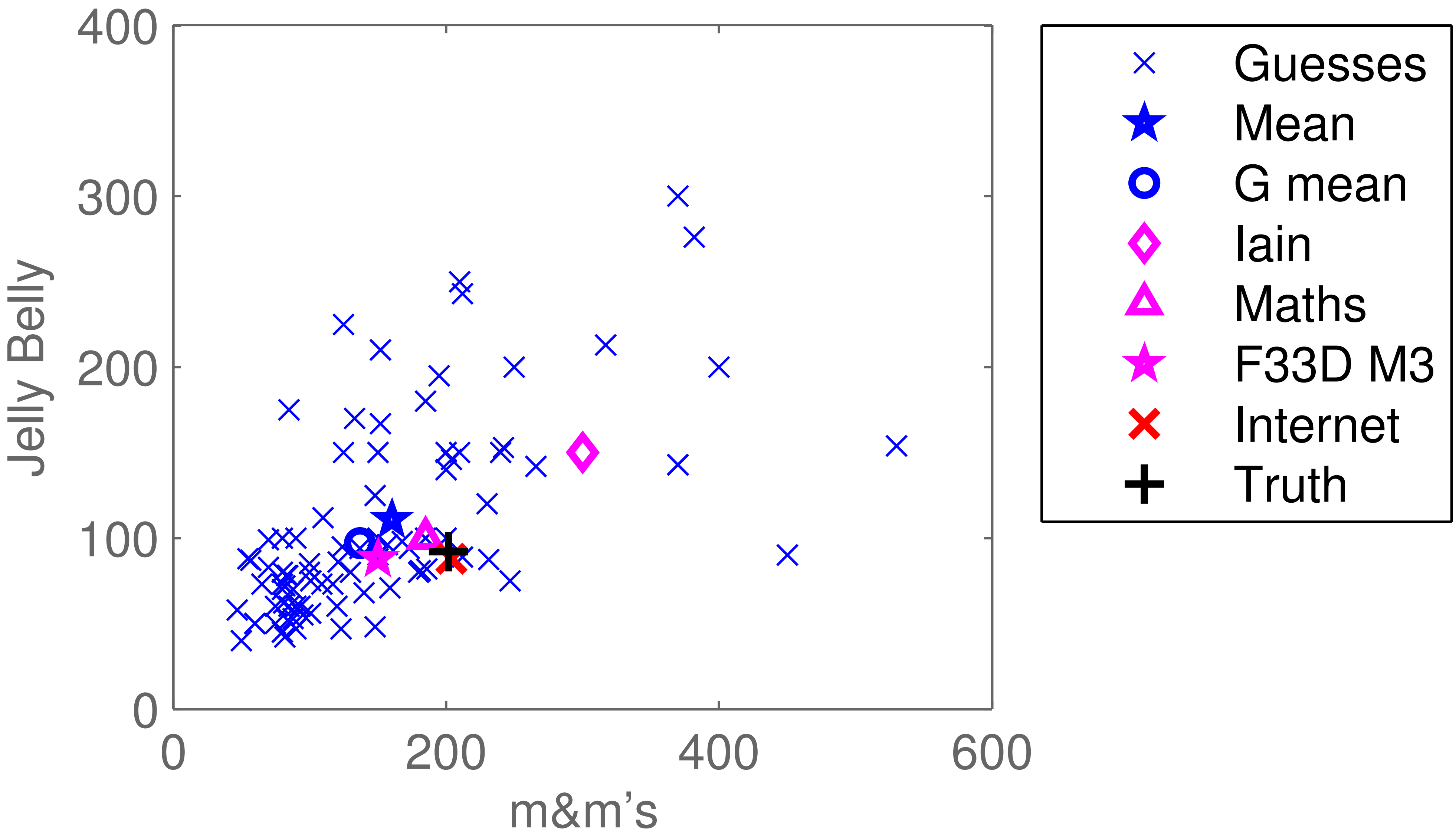
Number M&Ms: 231.25
Number Jelly Belly: 87.5
Num. choc-raisin blobs: 133.34
Full name: ANON
(to award prize only)

$\rho = 1\frac{g}{cm^3}$, $\rho = .7\frac{g}{cm^3}$

$.5cm^3$ each

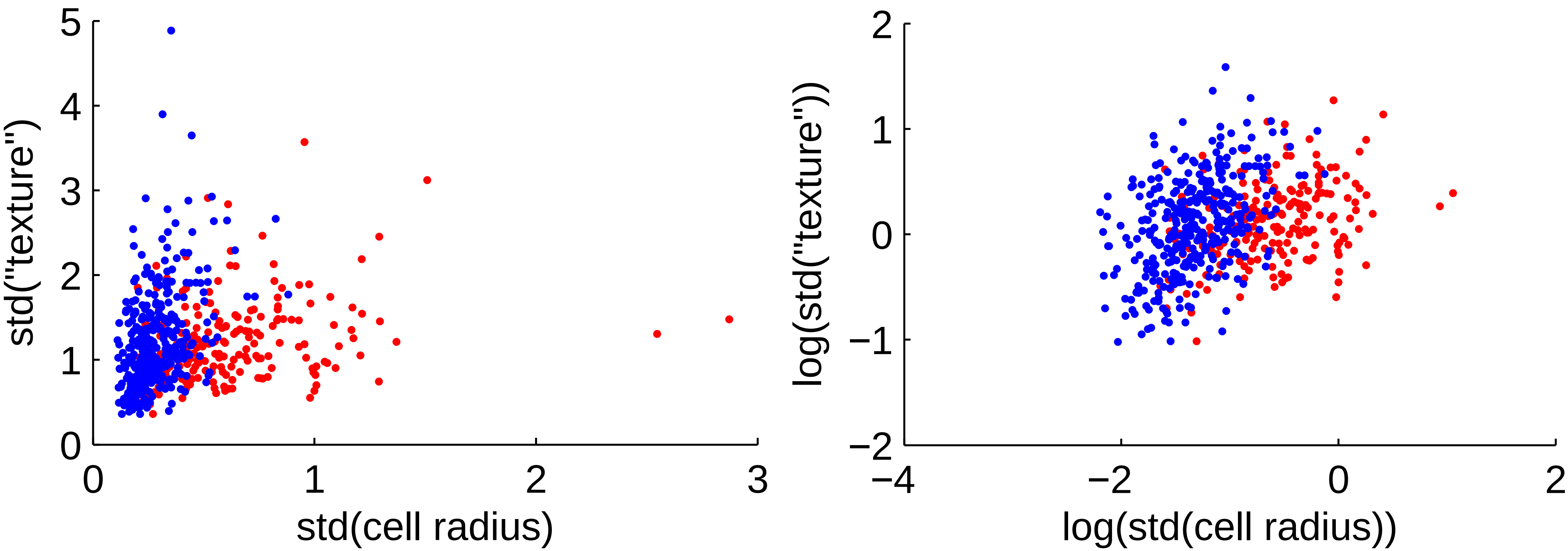
$\rho = \frac{m}{V} \Rightarrow m = \rho V$

A 2D space



For 3D and more, check out the code on the website.

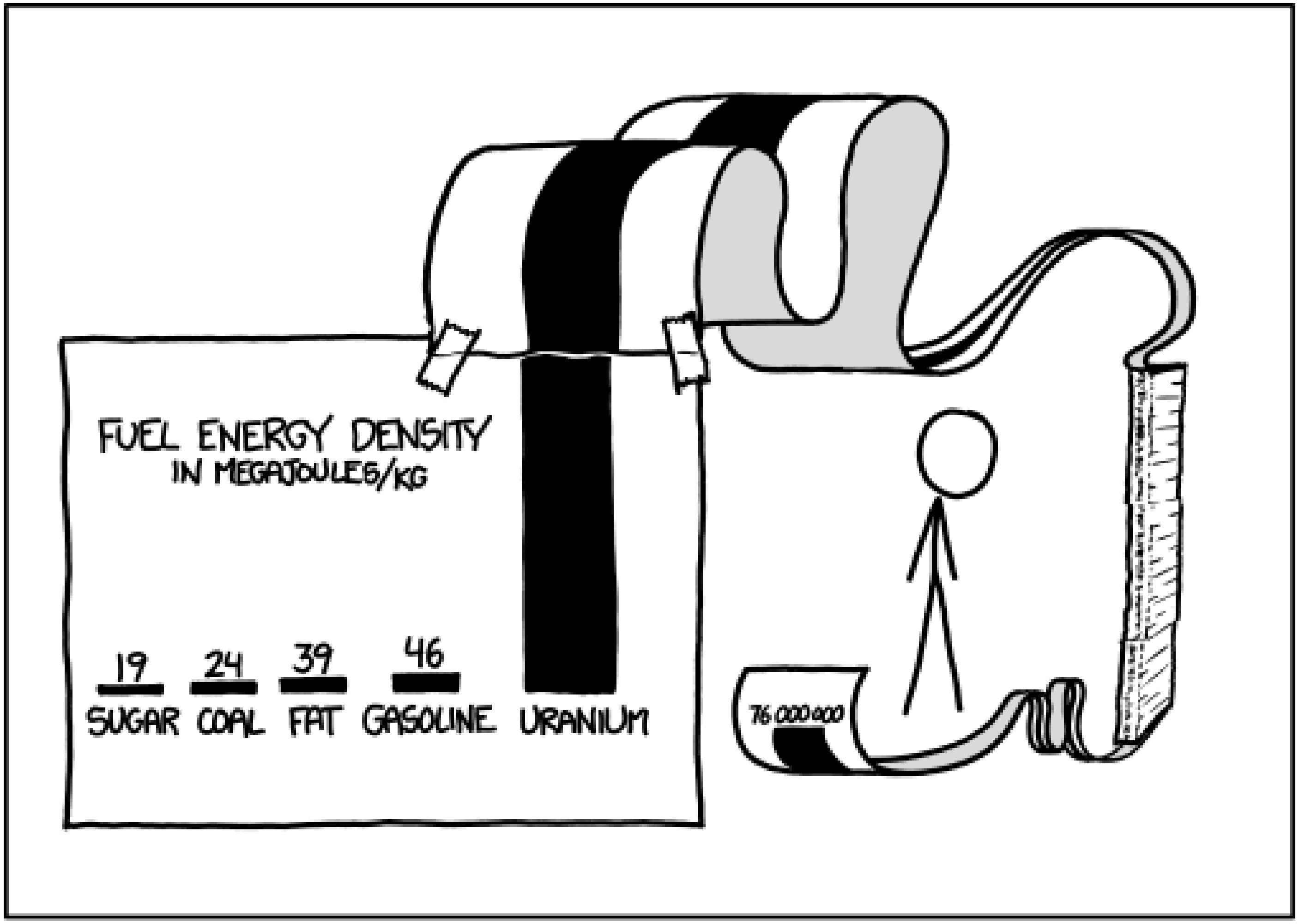
Often log-transform +ve data



Wisconsin breast cancer data

UCI ML repository

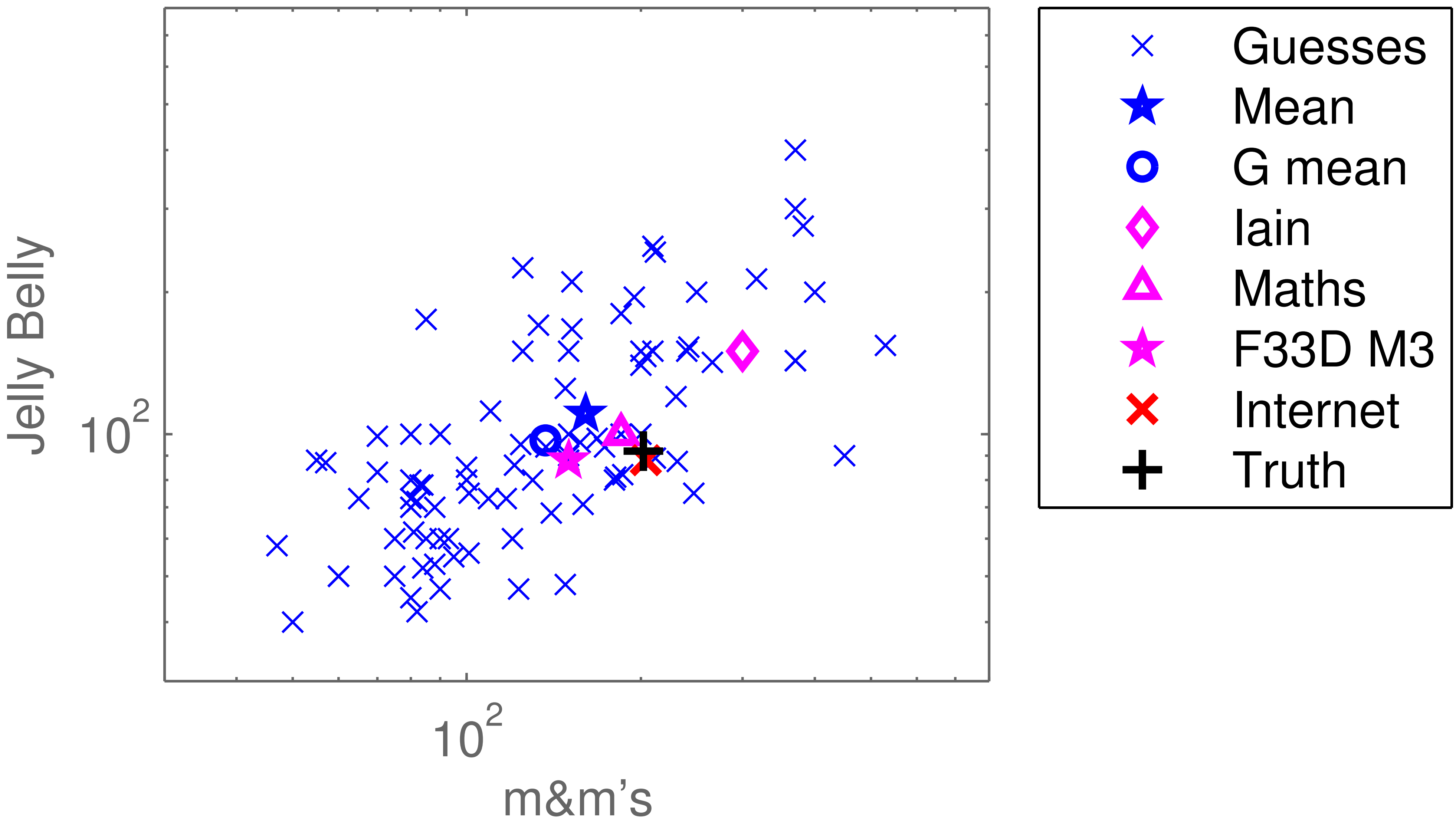
On taking logs



SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T FIND ENOUGH PAPER TO MAKE THEIR POINT *PROPERLY.*

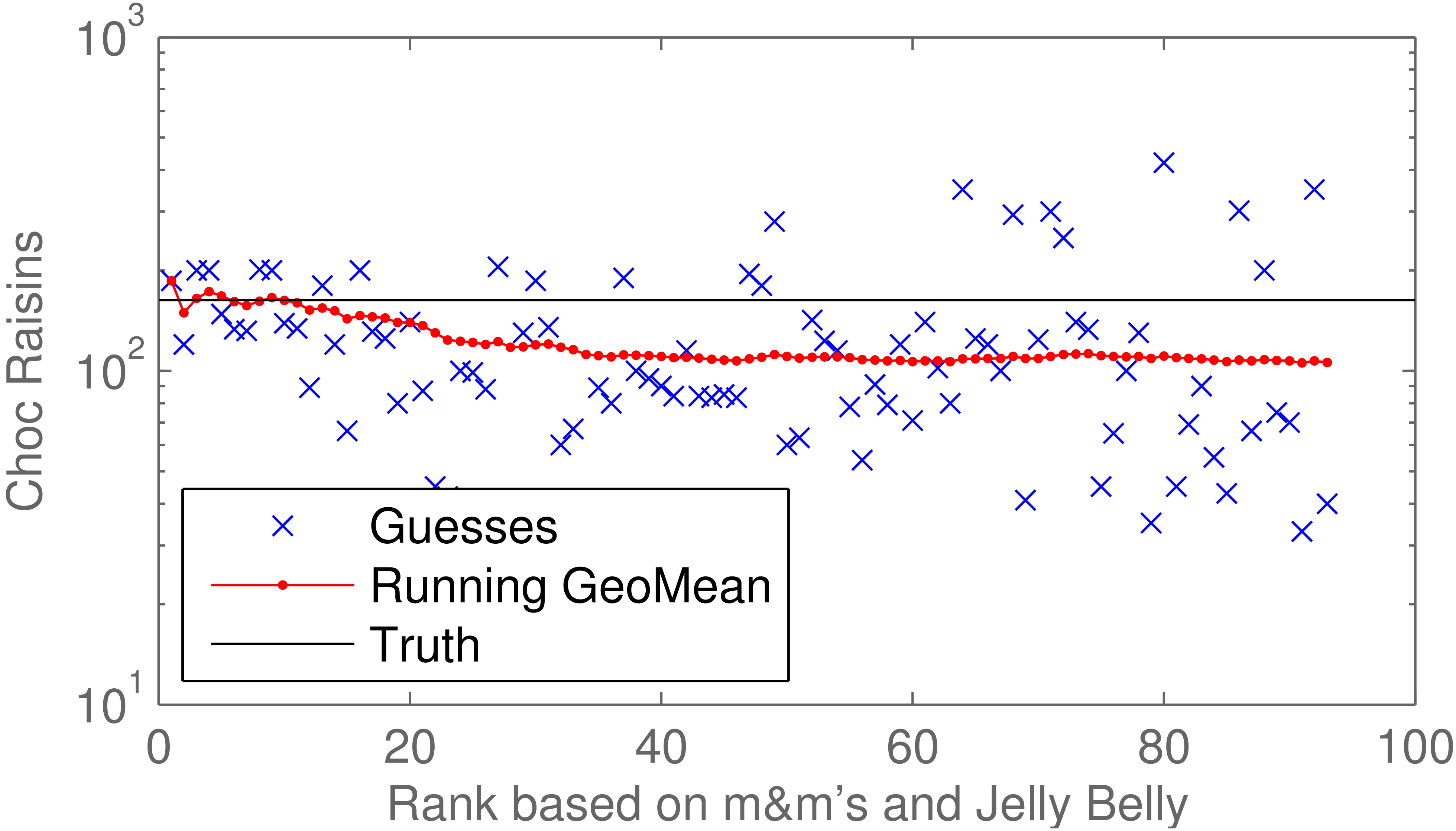
http://xkcd.com/1162/

Count guesses on log-scale



Were some people just lucky?

Ranking by past performance



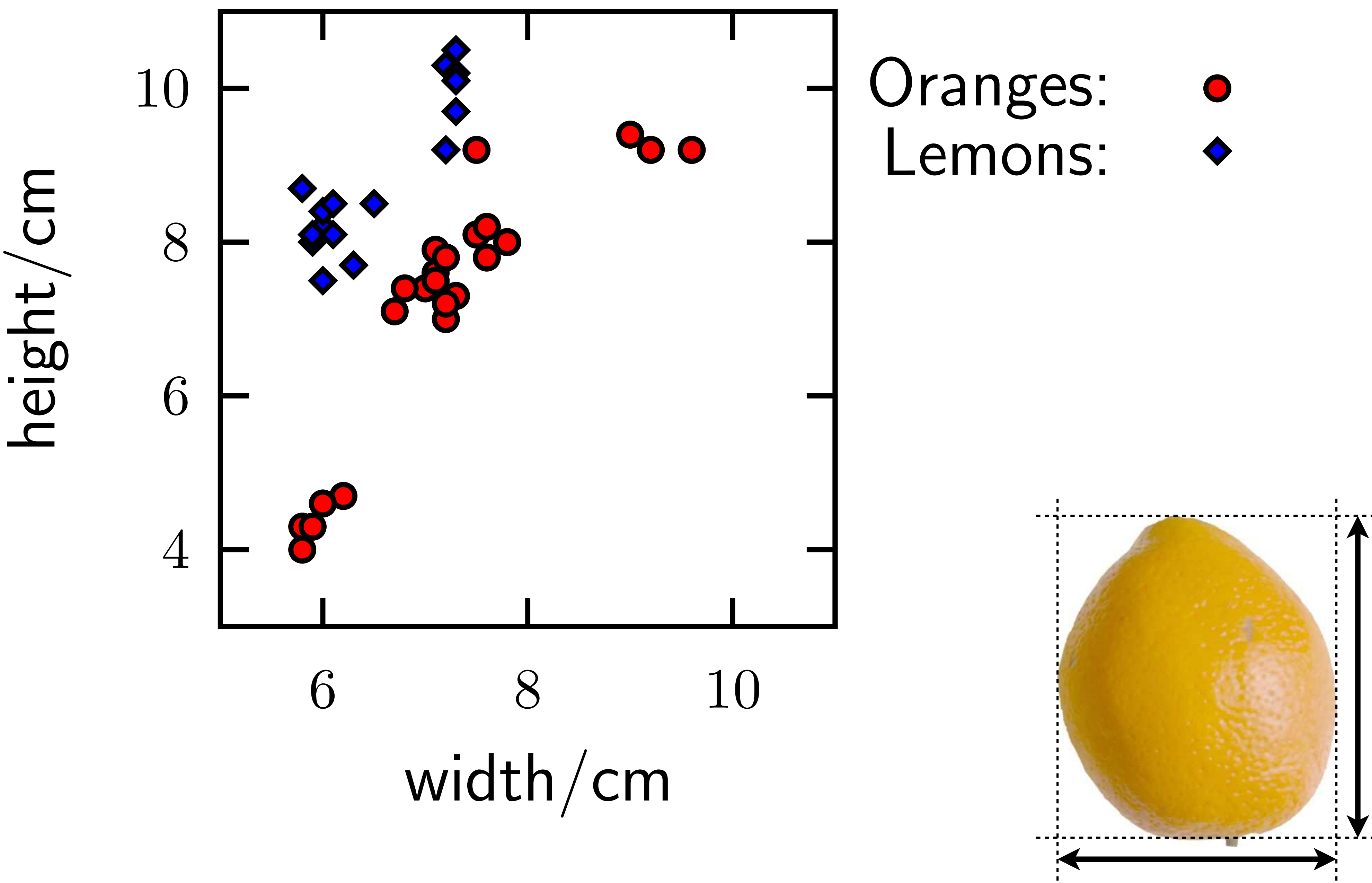
Today's Schedule:

— Collaborative counting (review)

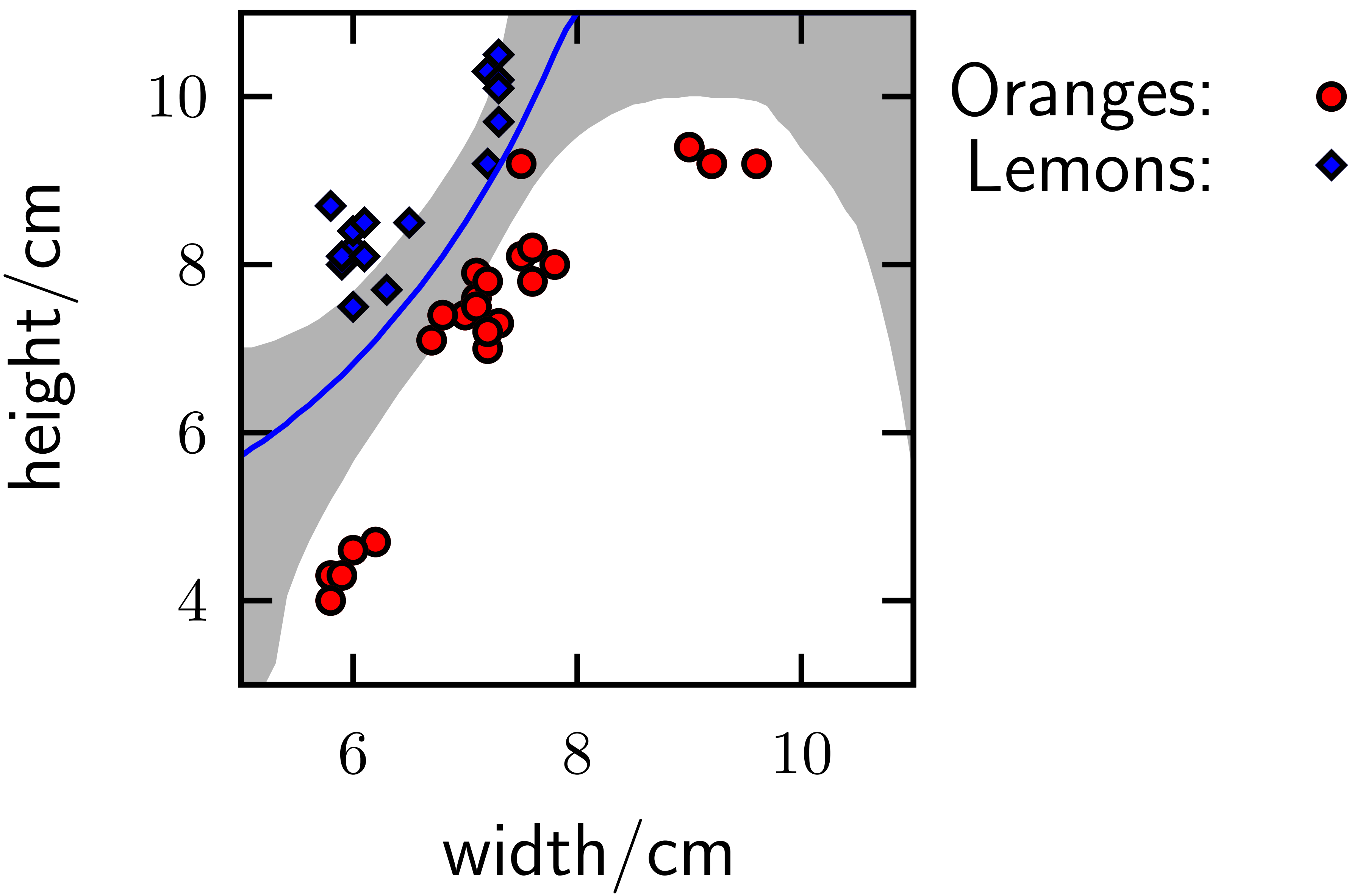
— **Clustering**

— How to stay on the road (time allowing)

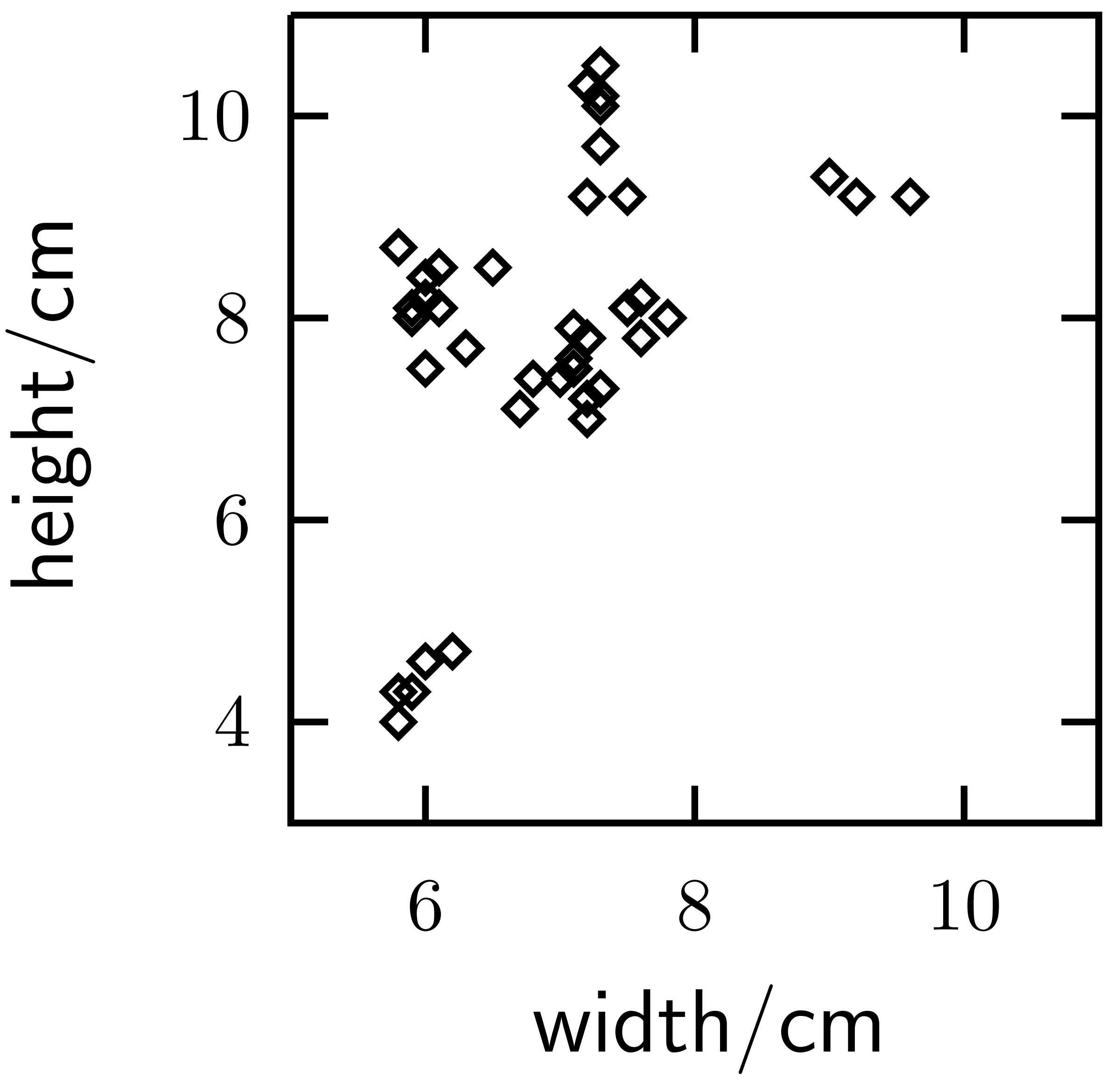
A two-dimensional space



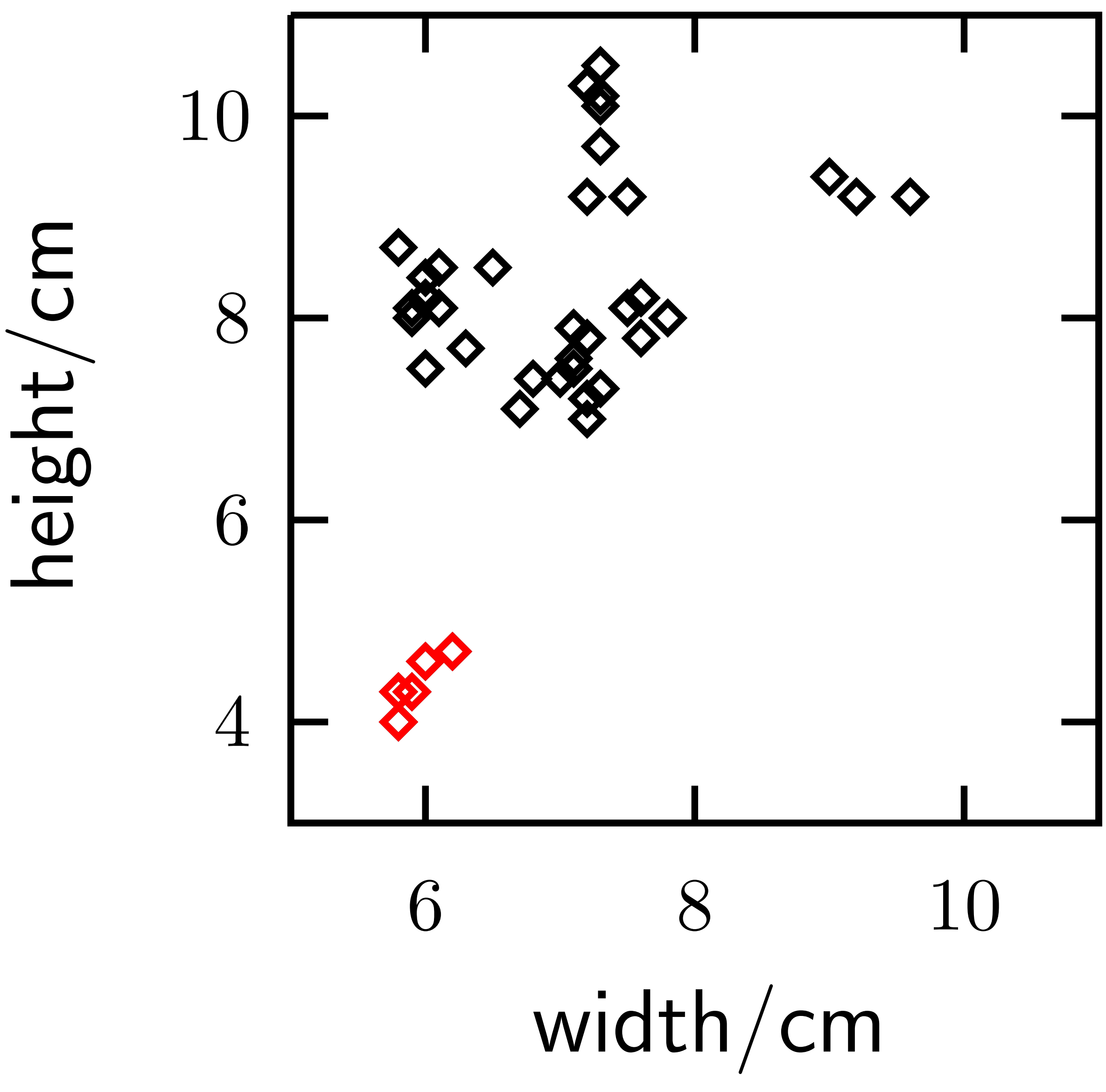
Supervised learning



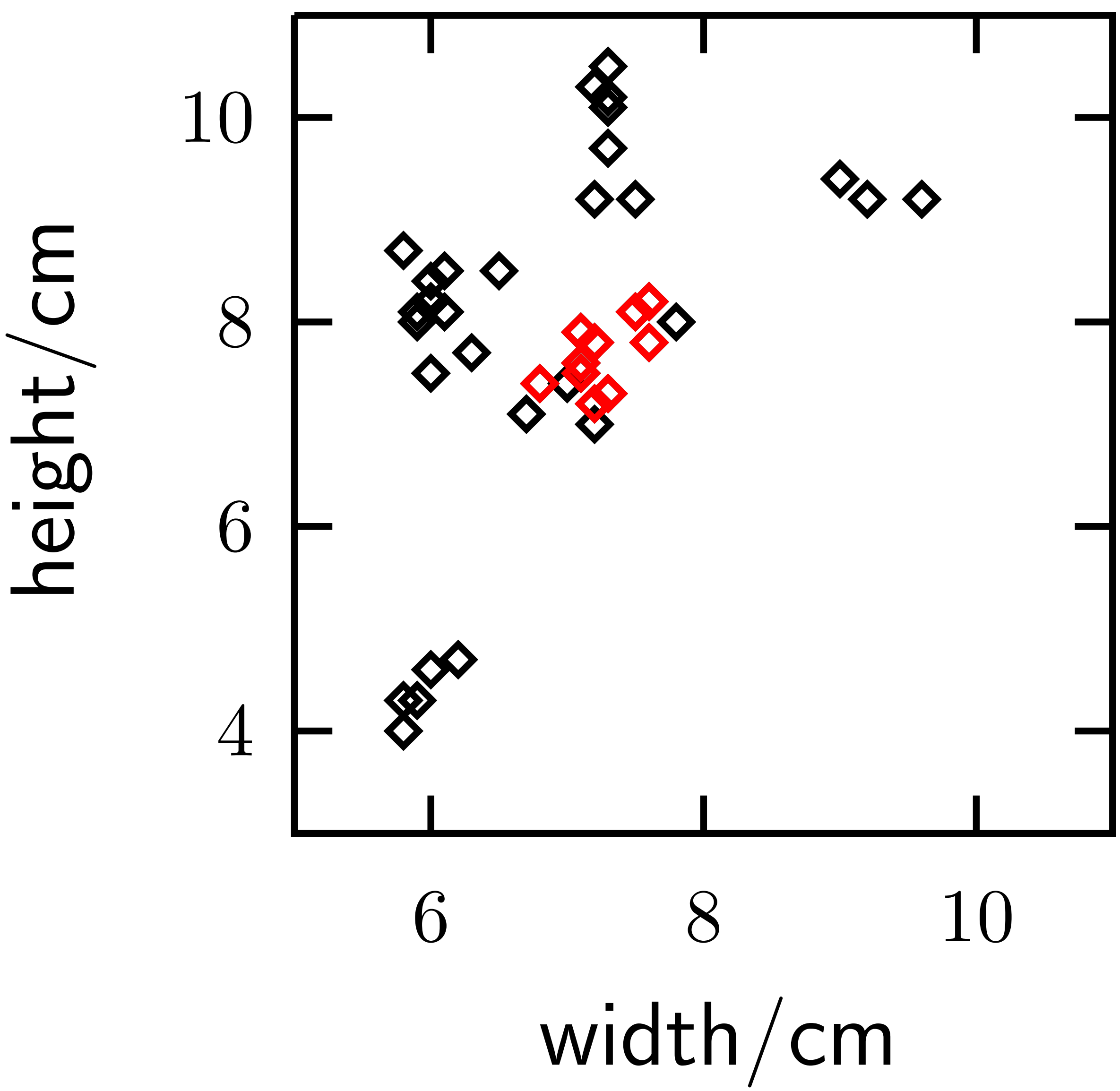
The Unsupervised data



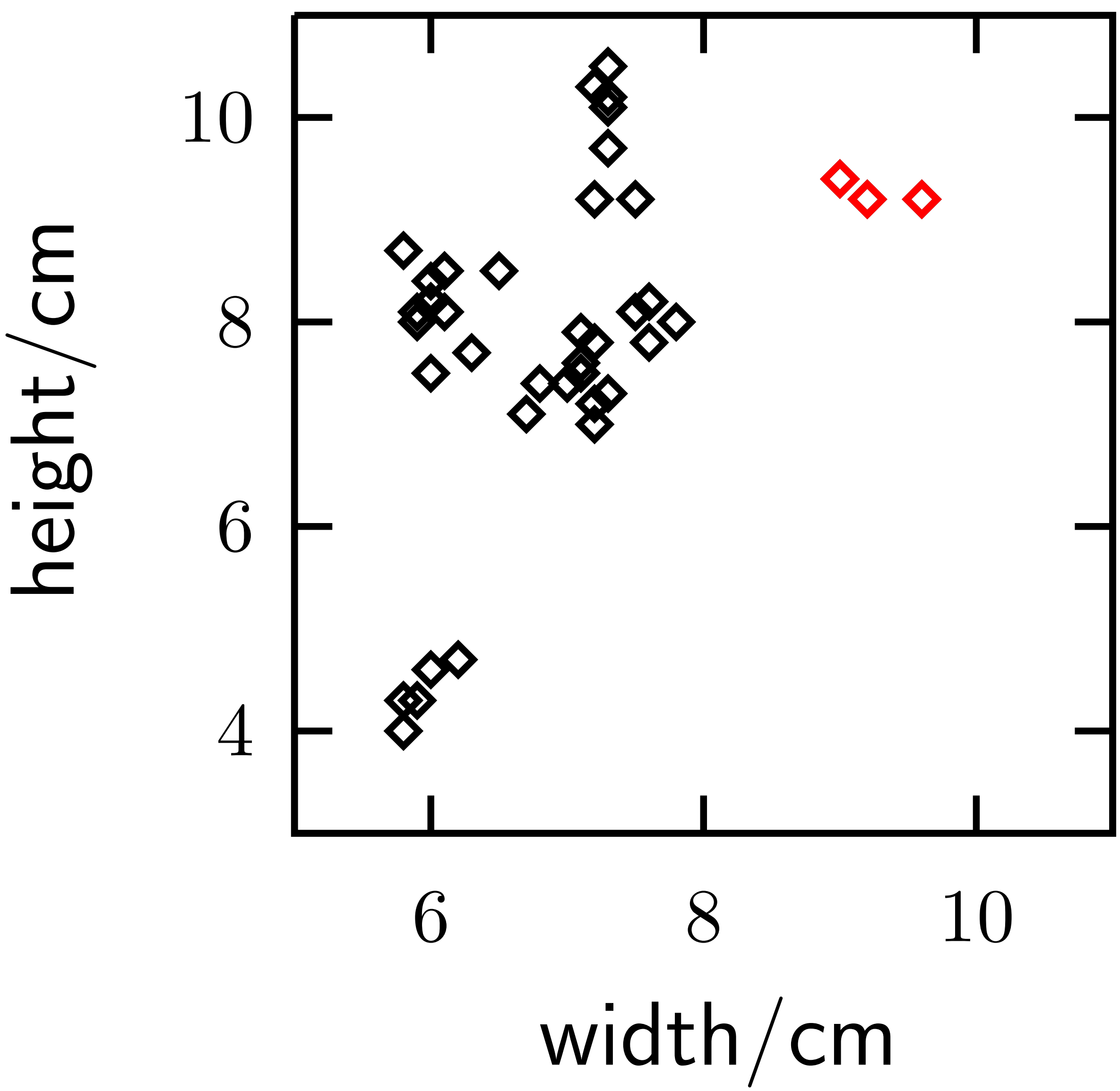
Manderins



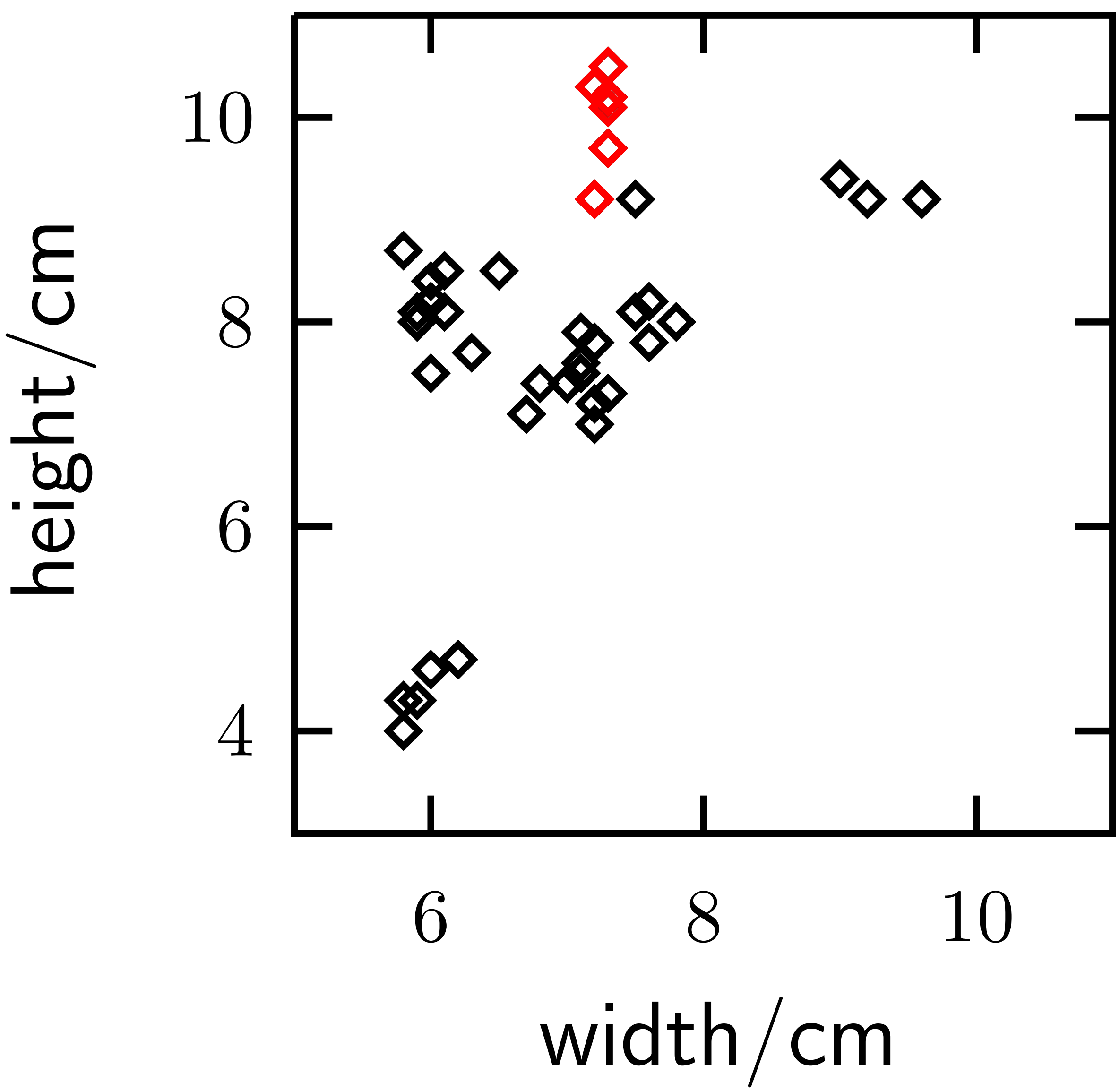
Navel oranges



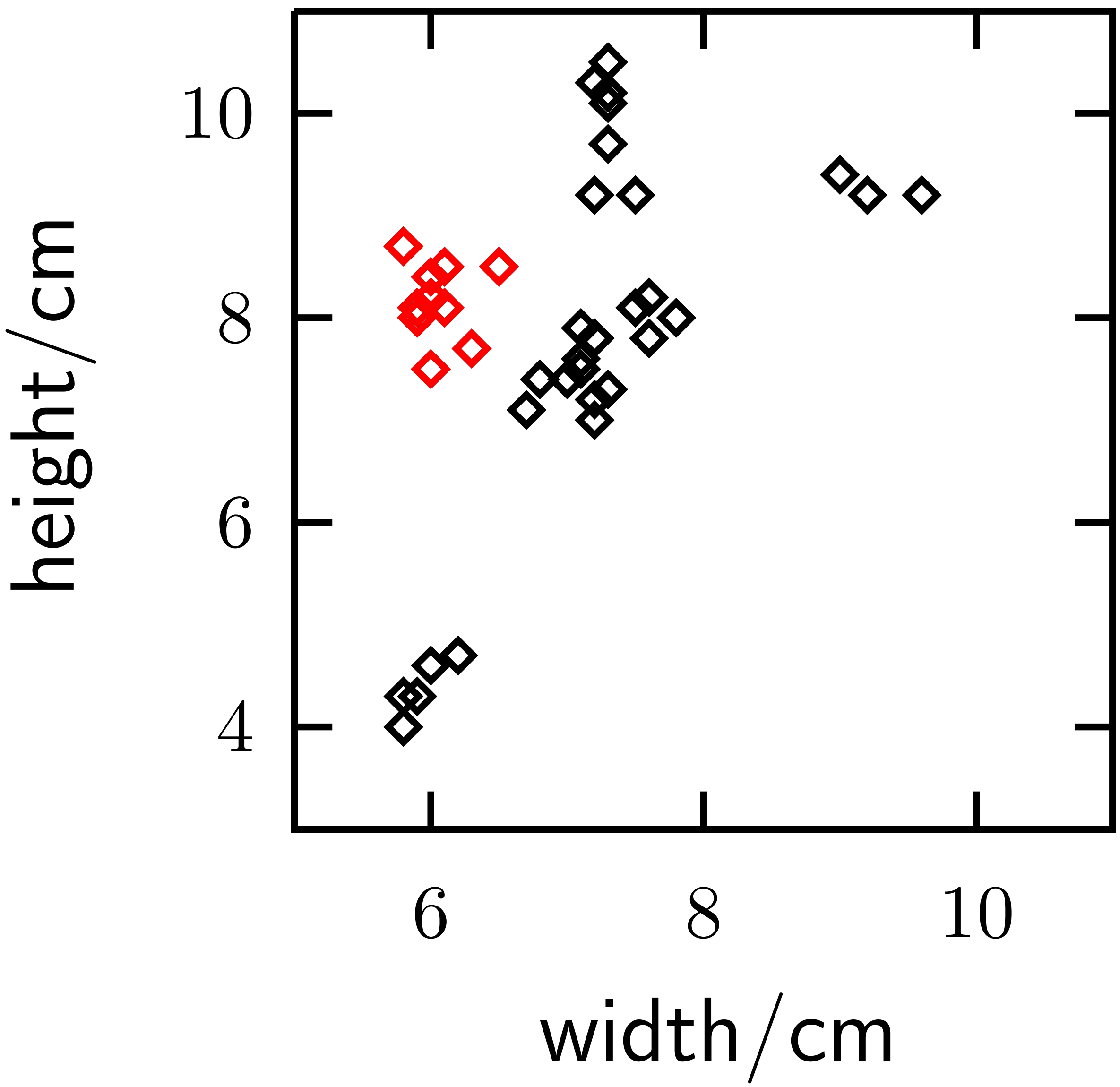
Spanish jumbo oranges



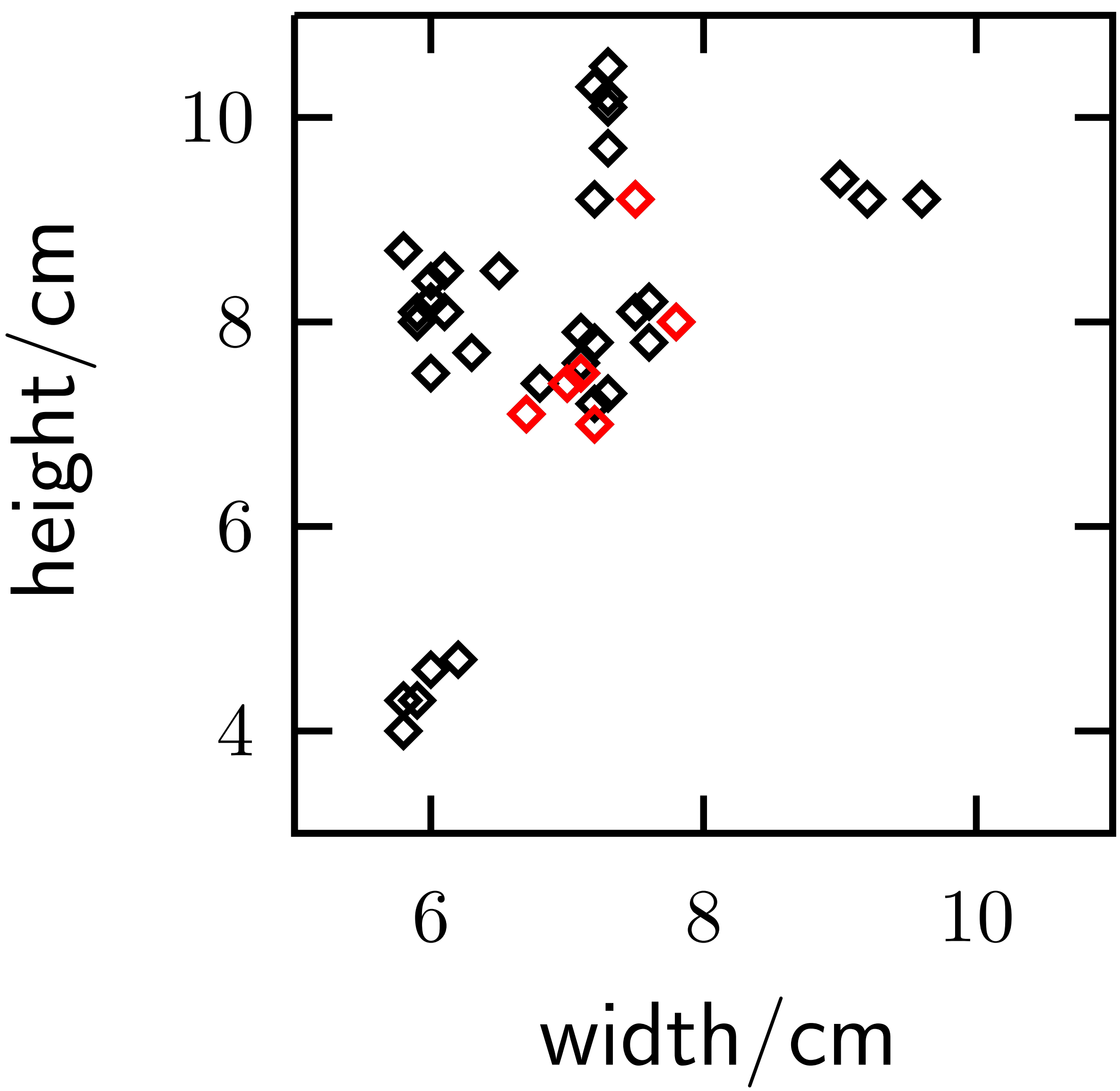
Belsan lemons



Some other lemons



“Seconds” Oranges



Clustering

"Human brains are good at finding regularities in data. One way of expressing regularity is to put a set of objects into groups that are similar to each other. For example, biologists have found that most objects in the natural world fall into one of two categories: things that are brown and run away, and things that are green and don't run away. The first group they call animals, and the second, plants."

Recommended reading: David MacKay textbook, p284–

`http://www.inference.phy.cam.ac.uk/mackay/itila/`

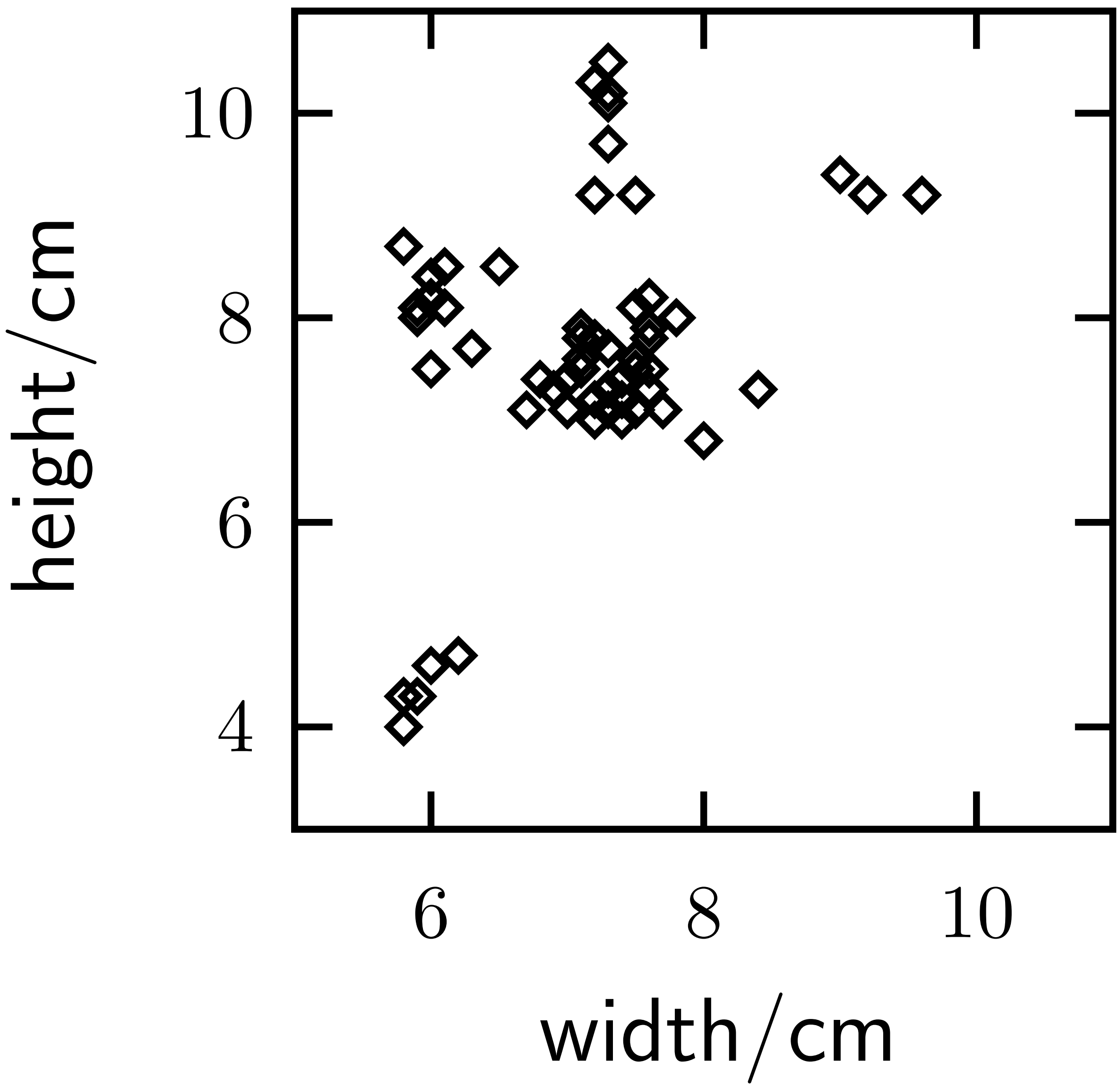
K-means clustering

A simple algorithm to find clusters:

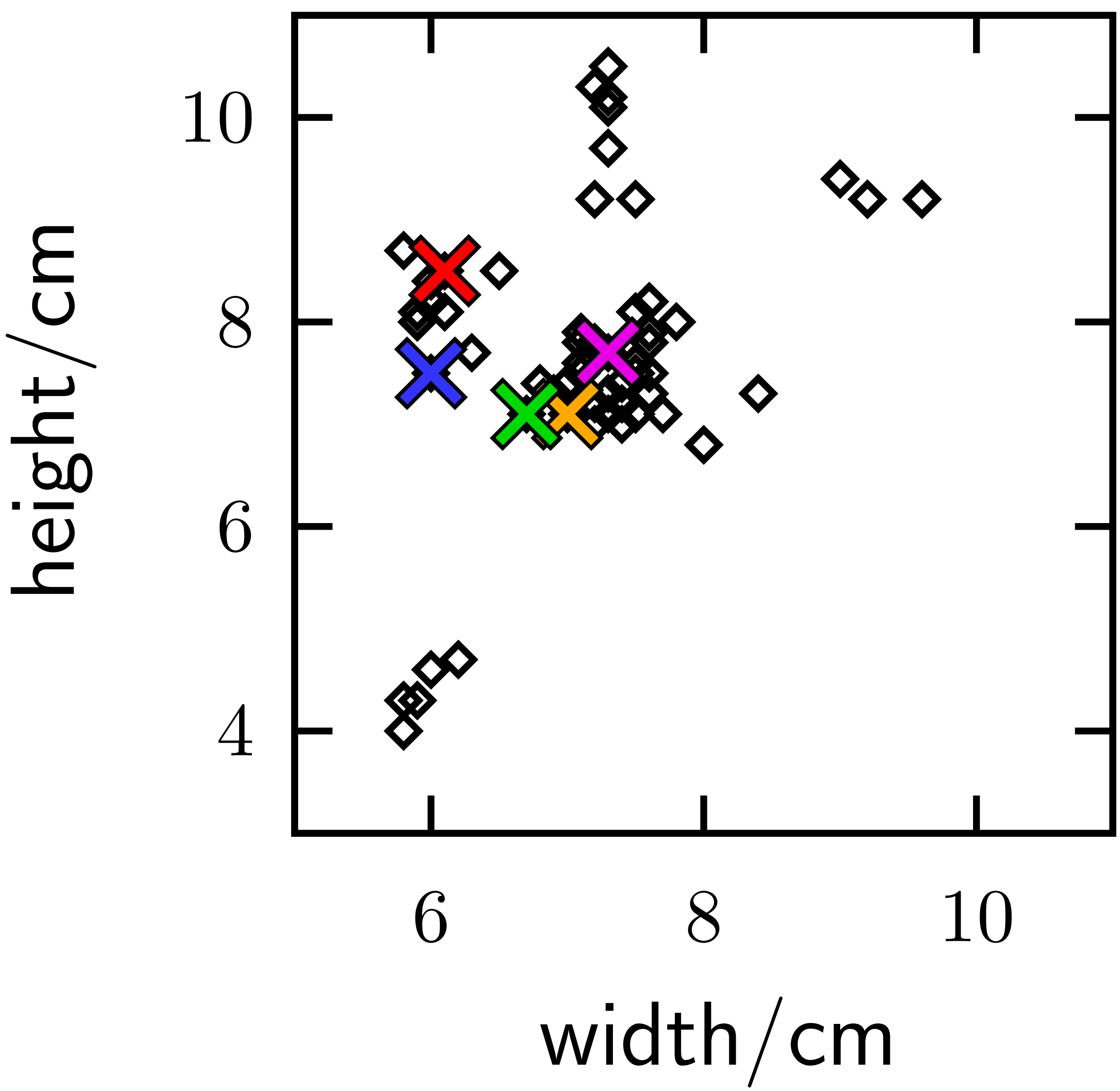
1. Pick K random points as cluster center positions
2. Assign each point to its nearest center*
3. Move each center to mean of its assigned points
4. If centers moved, `goto` 2.

* In the unlikely event of a tie, break tie in some way.
For example, assign to the center with smallest index in memory.

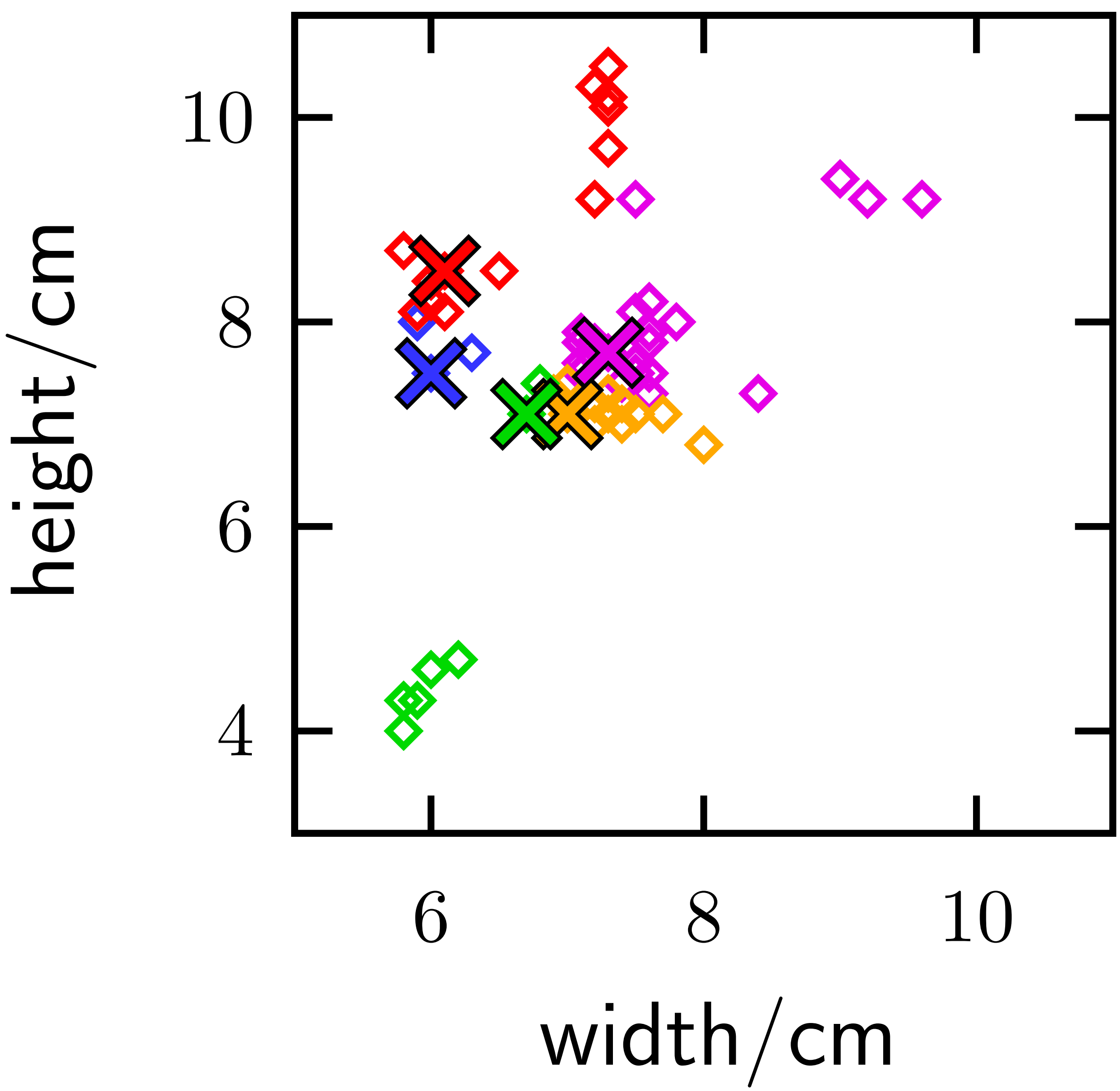
K-means clustering



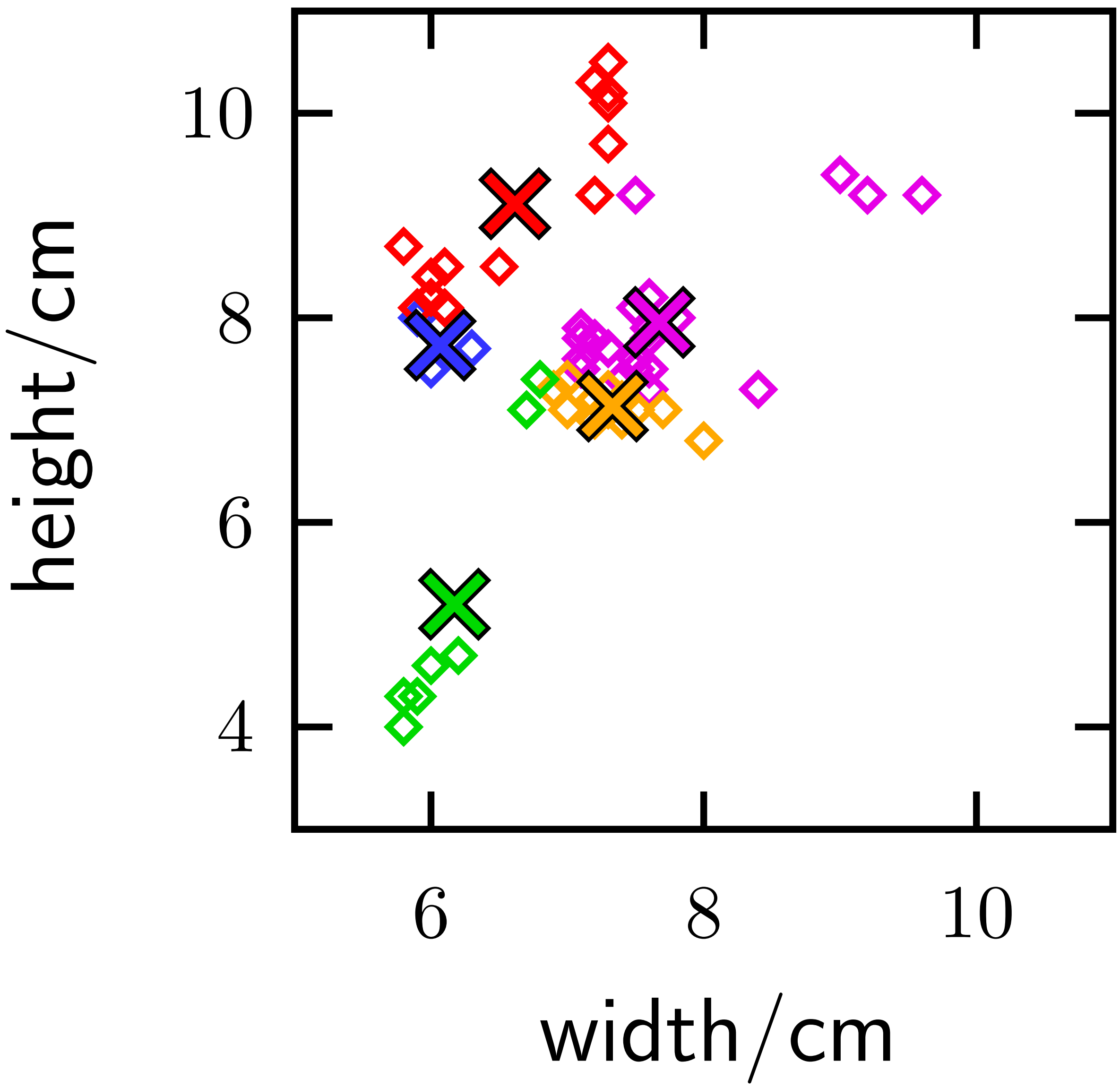
K-means clustering



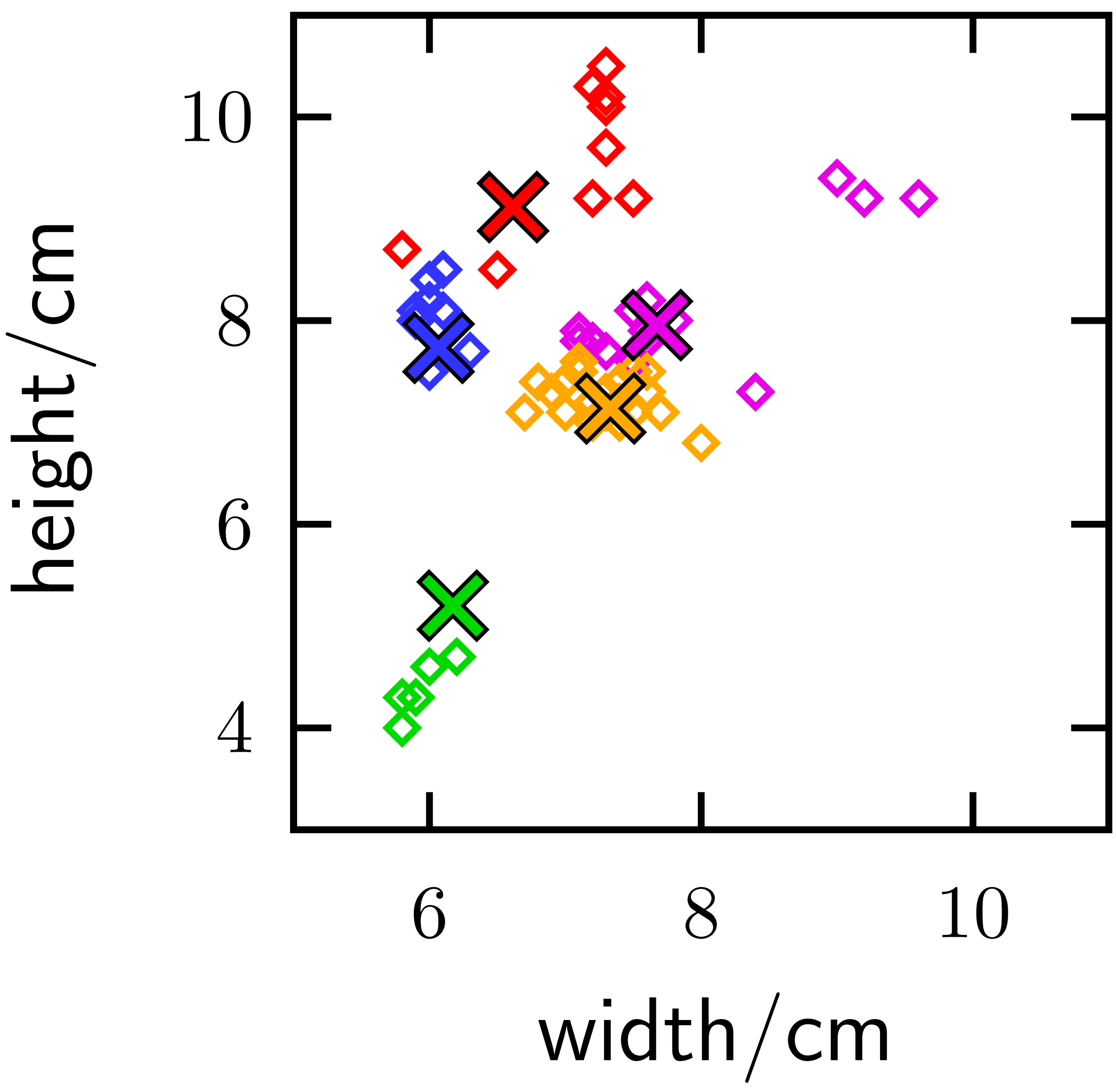
K-means clustering



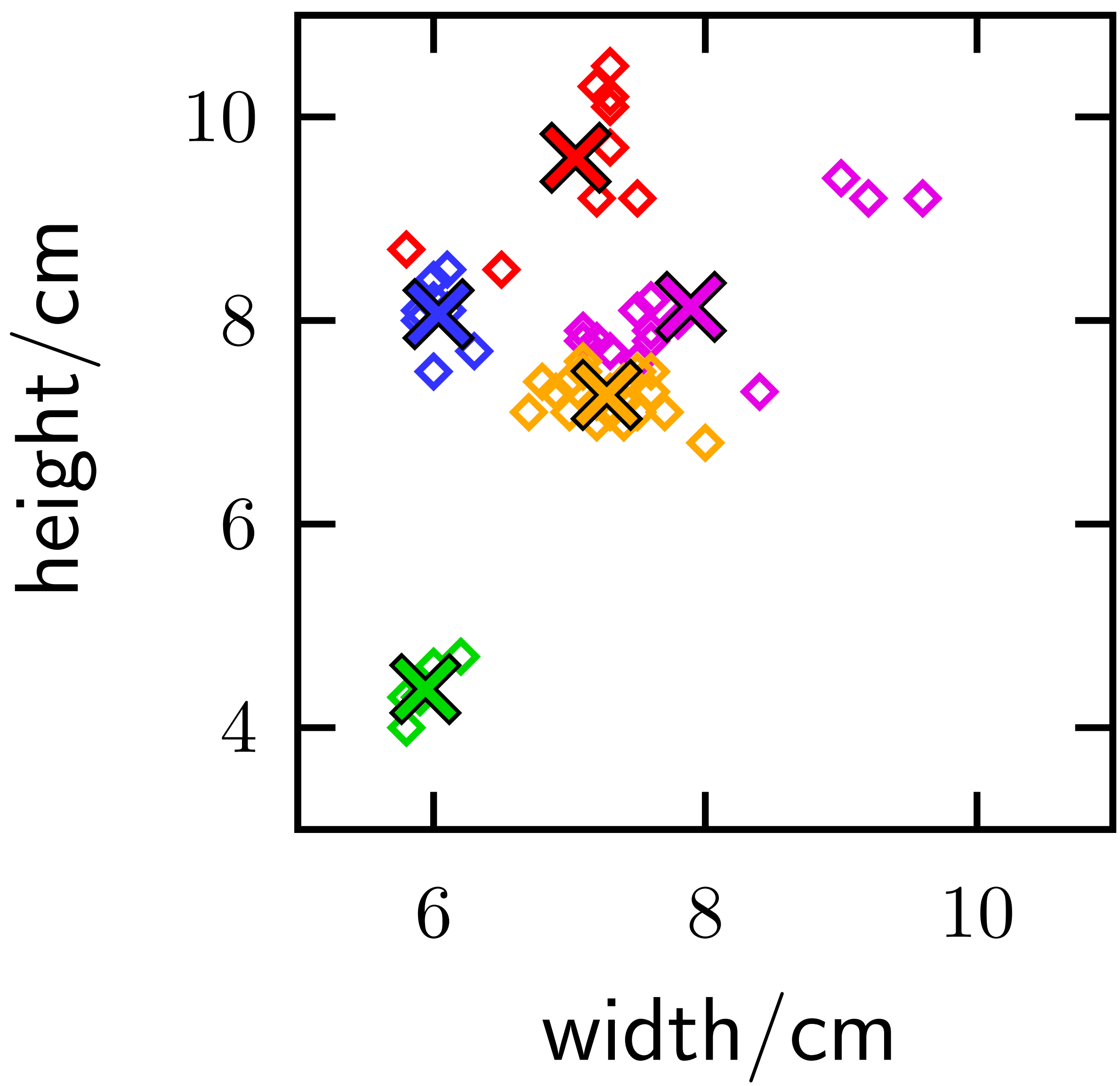
K-means clustering



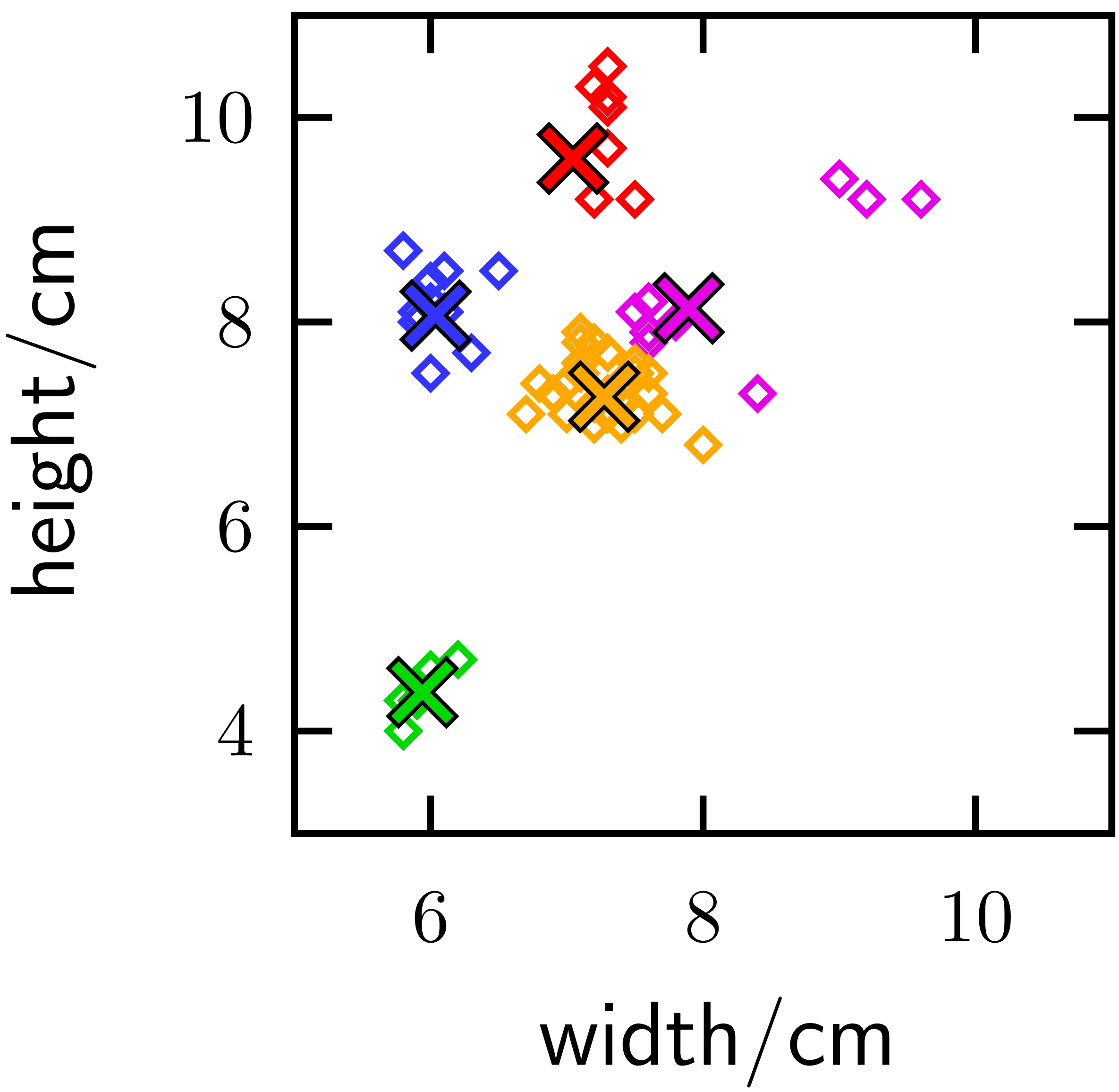
K-means clustering



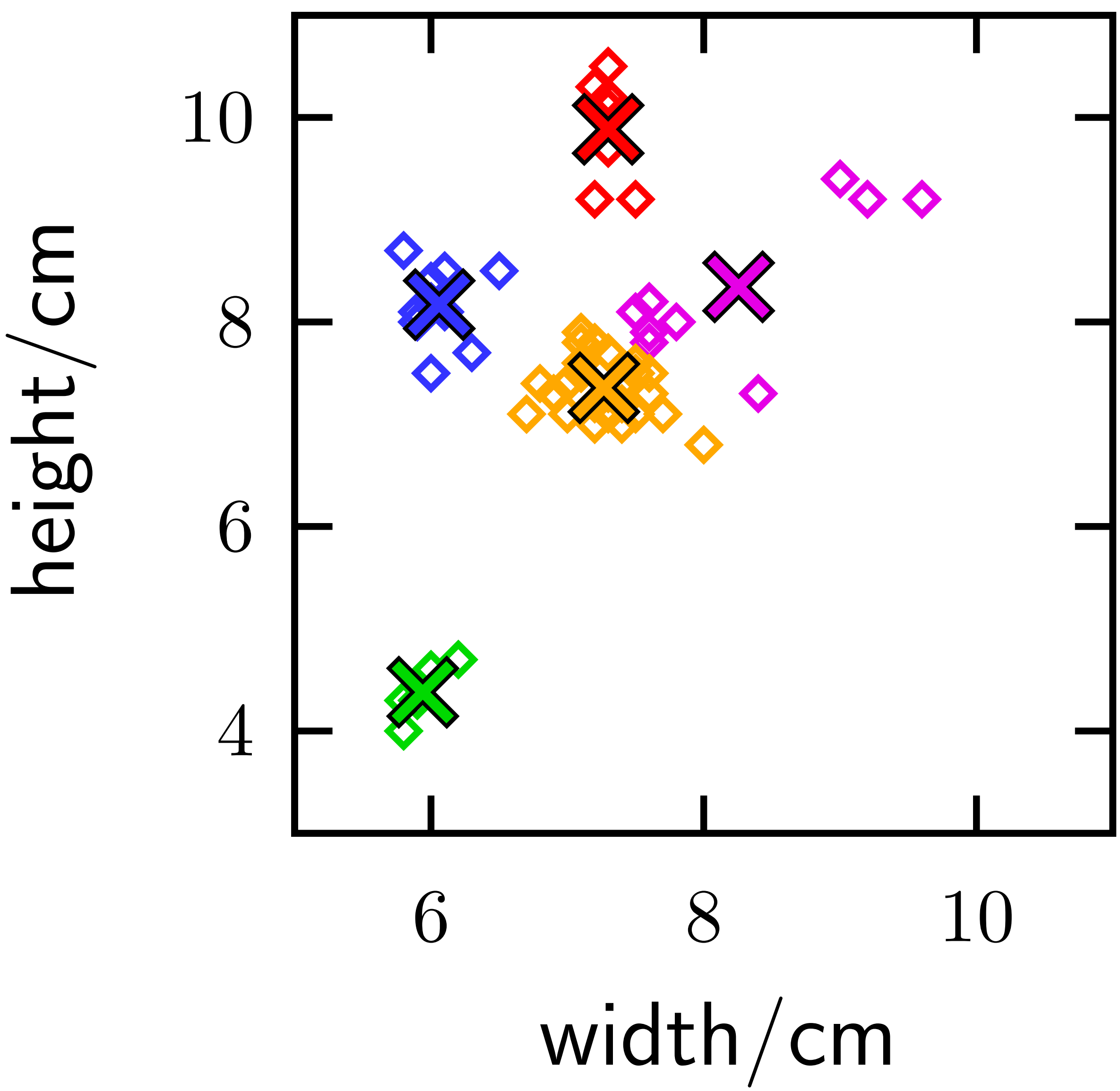
K-means clustering



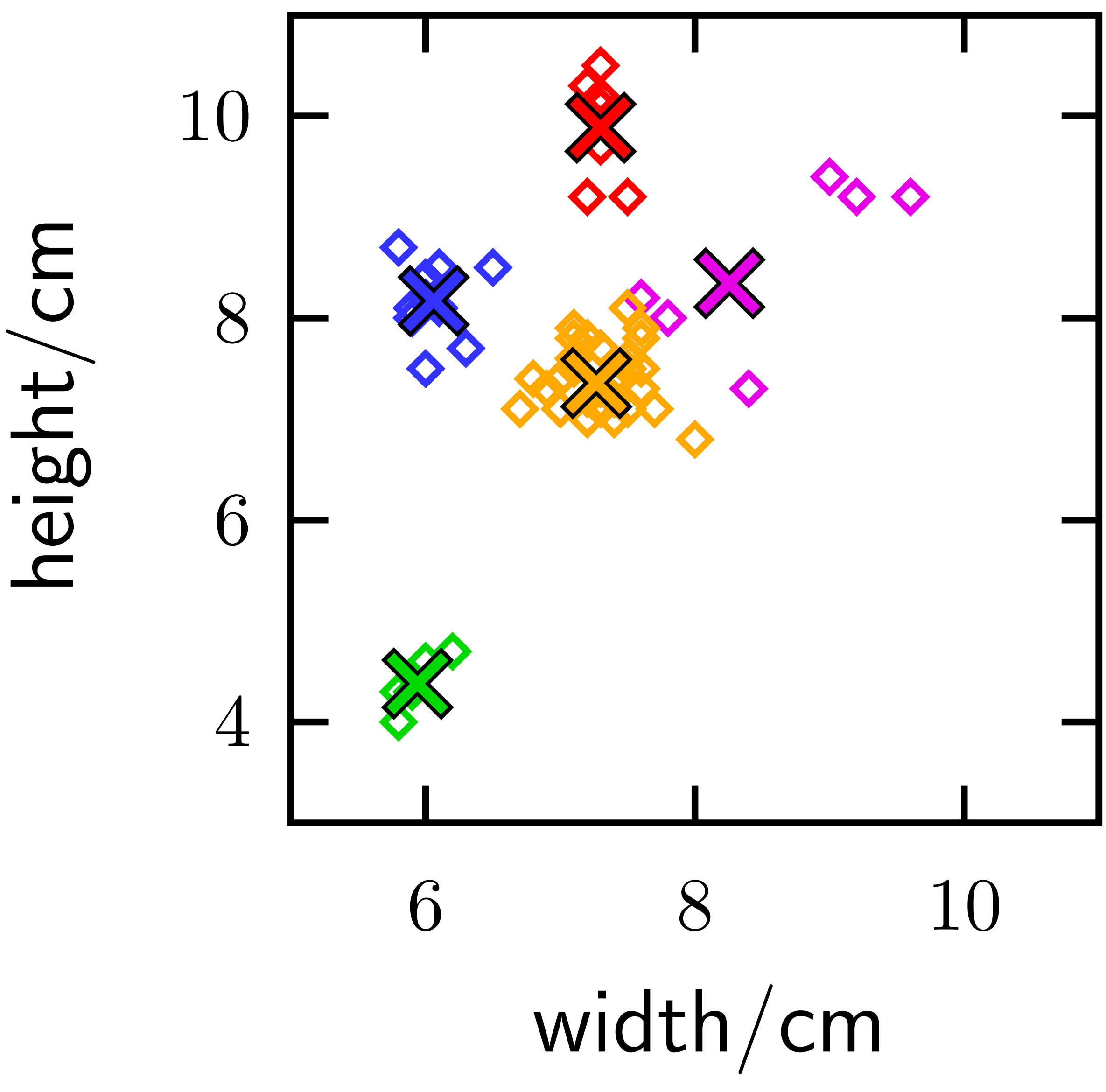
K-means clustering



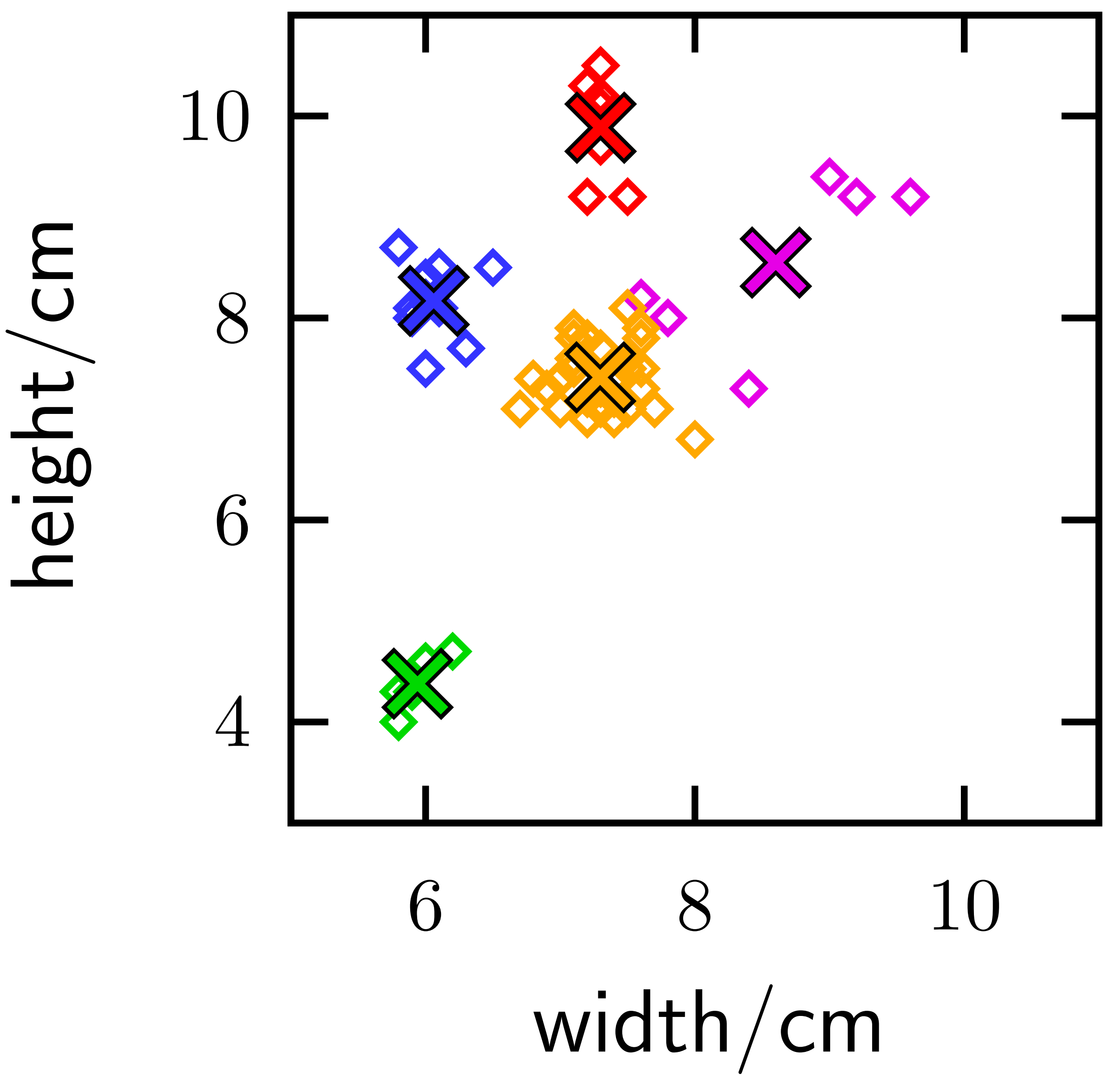
K-means clustering



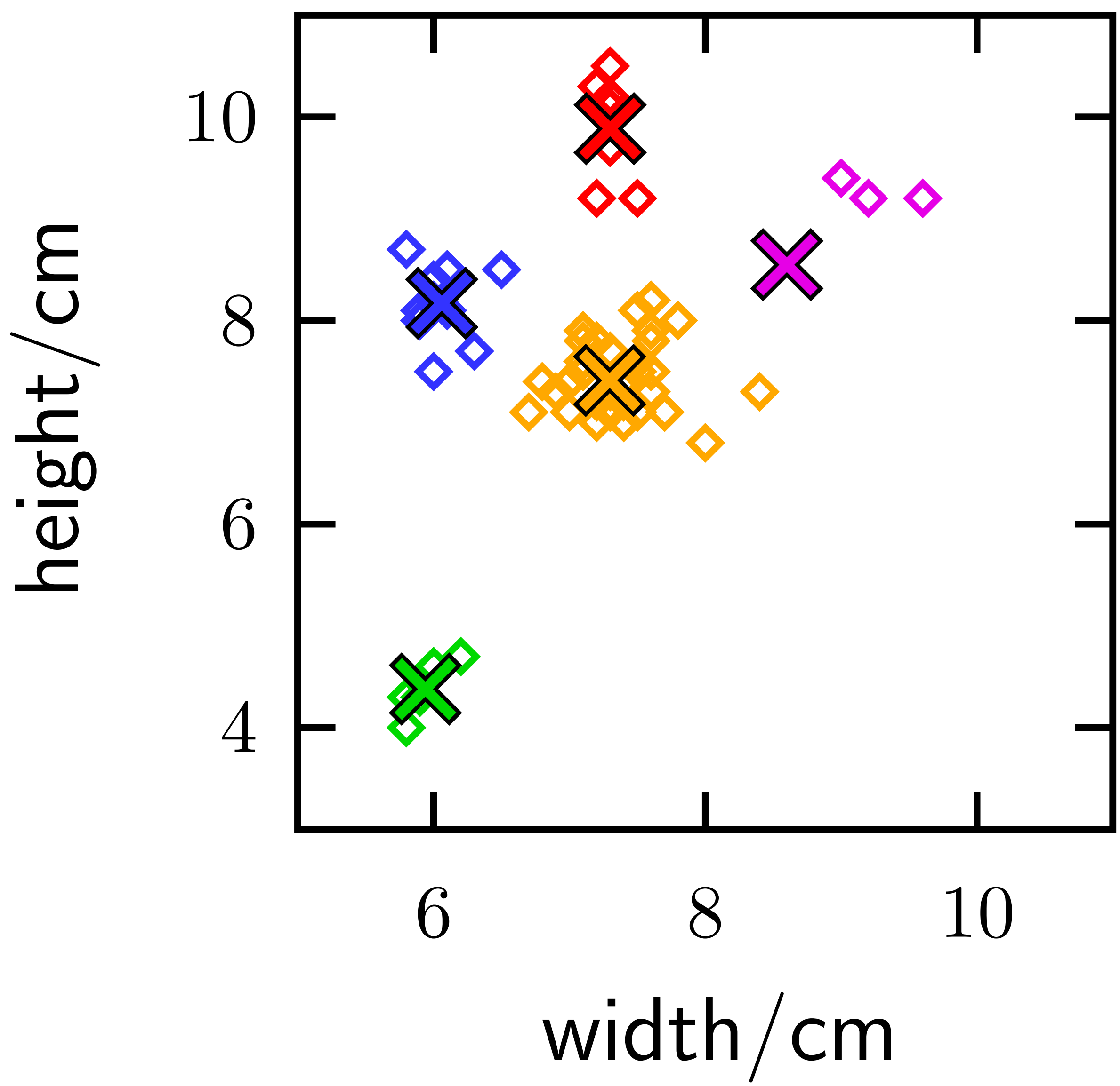
K-means clustering



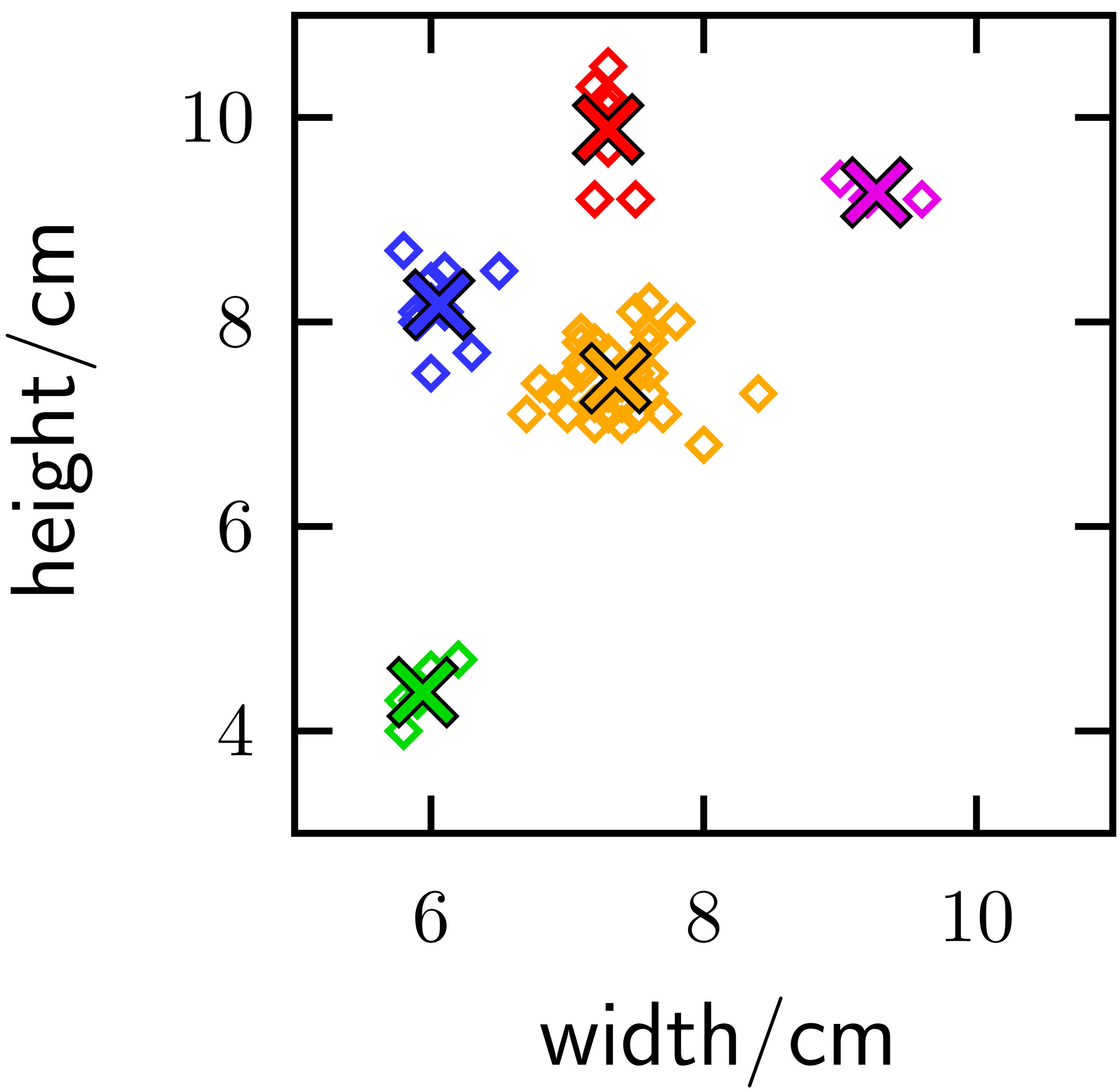
K-means clustering



K-means clustering



K-means clustering



Theory of K-means

If assignments don't change, algorithm terminates.

Can assignments cycle, never terminating?

Convergence proof technique: find a *Lyapunov function* $\mathcal{L}$, that is bounded below and cannot increase.

$\mathcal{L} =$ sum of square distances between points and centers

K**-means is an optimization algorithm** for $\mathcal{L}$.
Local optima are found. Running multiple times and using solution with best $\mathcal{L}$ is common.

Today's Schedule:

— Collaborative counting (review)

— Clustering

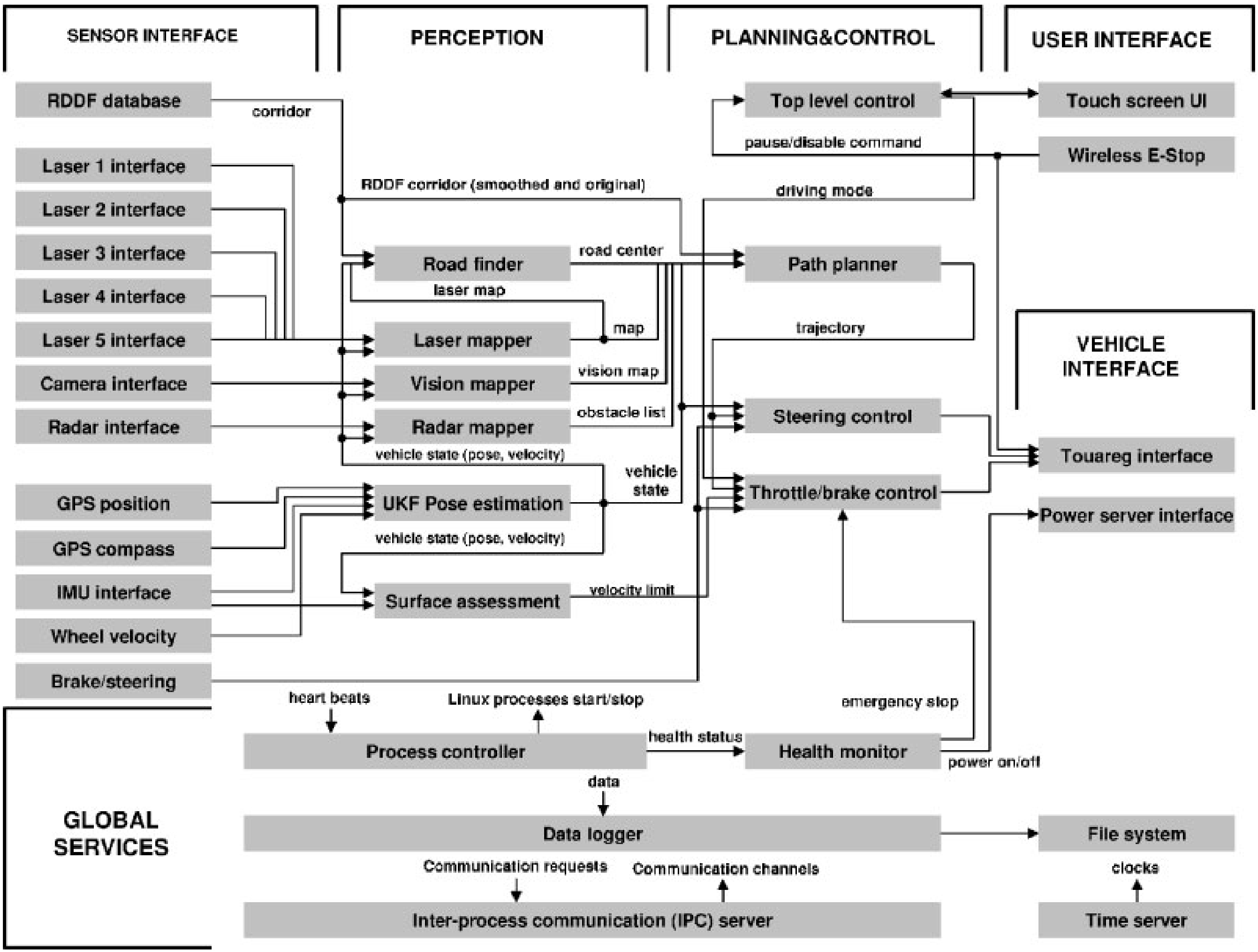
— **How to stay on the road** (time allowing)

Stanley

Stanford Raing Team; DARPA 2005 challenge

http://robots.stanford.edu/talks/stanley/

Inside Stanley

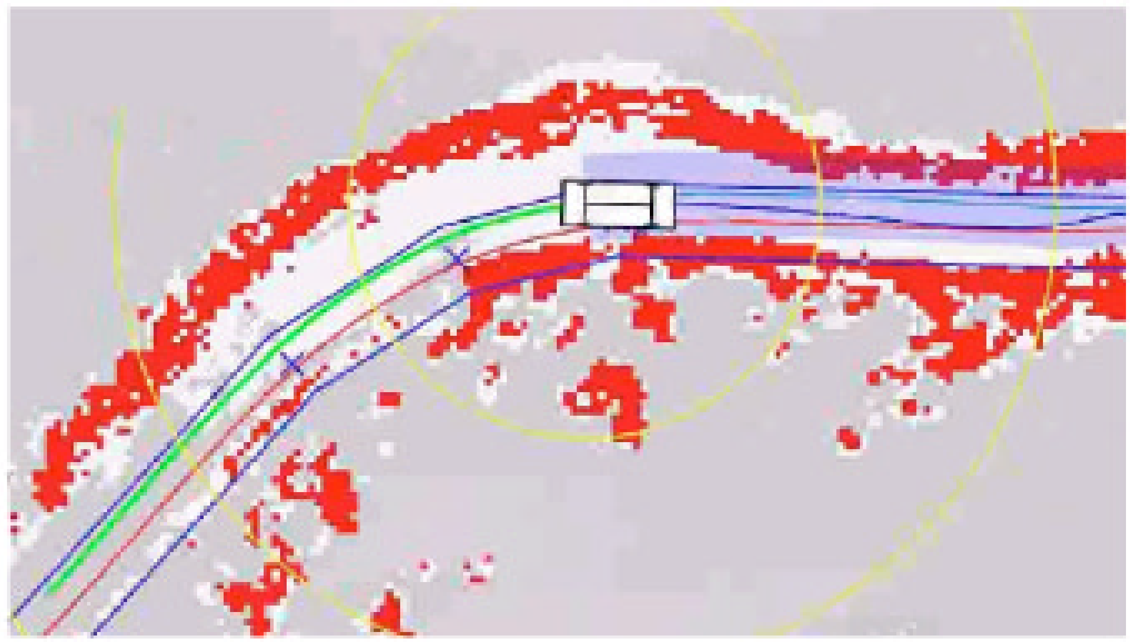


Stanley figures from Thrun et al., J. Field Robotics 23(9):661, 2006.

Perception and intelligence

(a) **Beer Bottle Pass**

(b) **Map and GPS corridor**

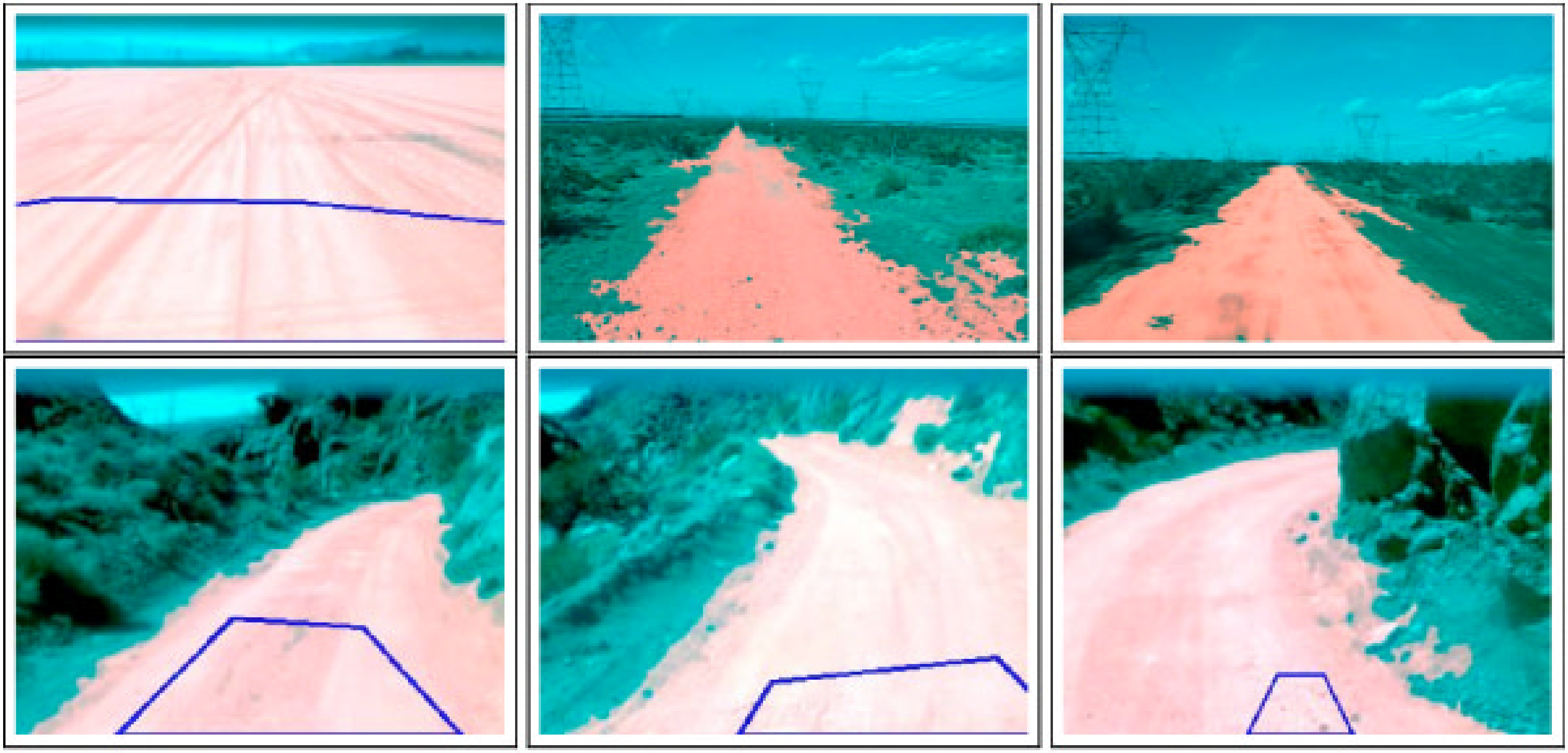


It would look pretty stupid to run off the road,
just because the trip planner said so.

How to stay on the road?

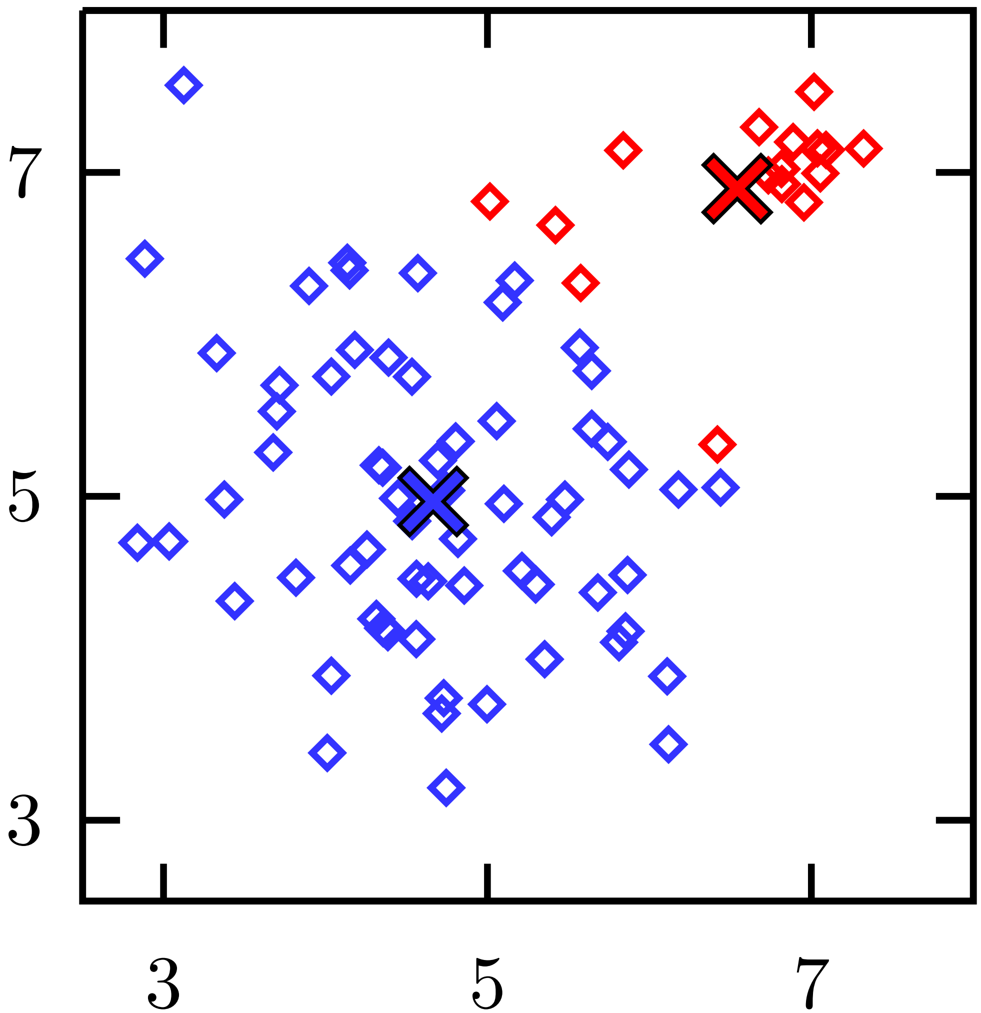
Classifying road seems hard. Colours and textures change: road appearance in one place may match ditches elsewhere.

Clustering to stay on the road



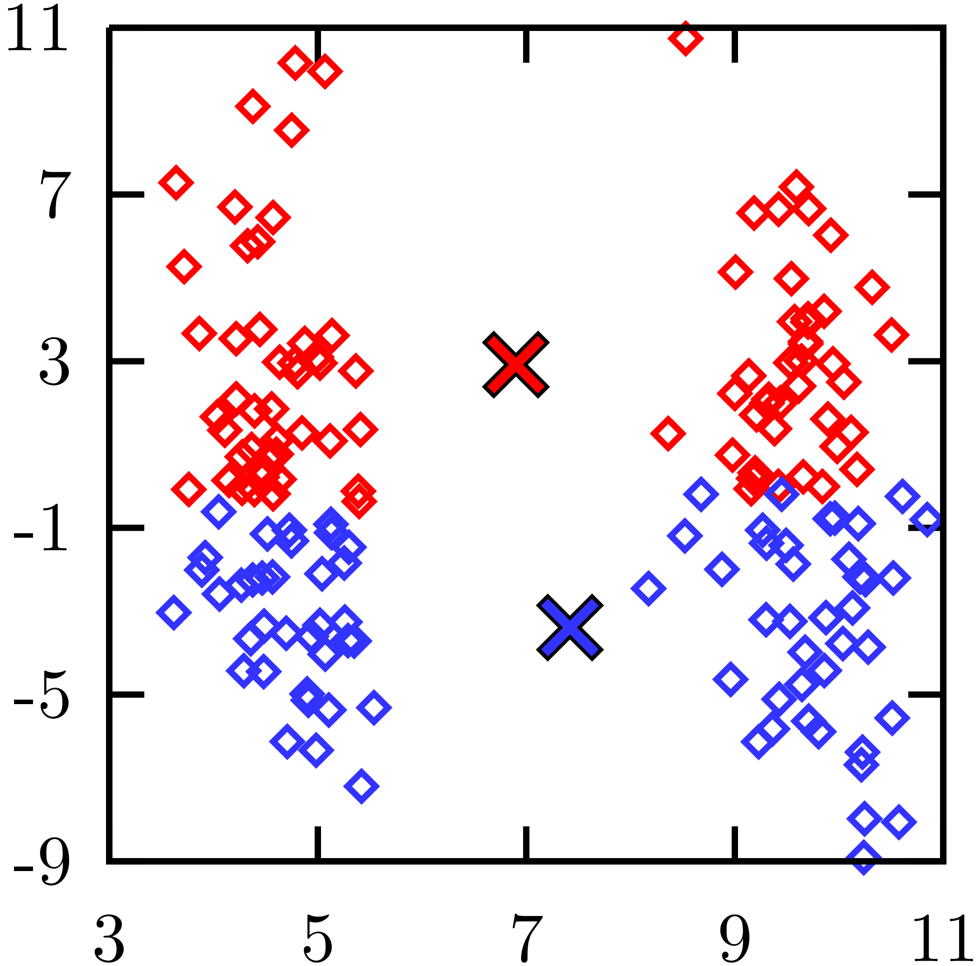
Stanley used a Gaussian mixture model. “Souped up K-means.”
The cluster just in front is road (unless we already failed).

Failures of K-means



Large clouds pull small clusters off-center

Failures of K-means



Distance needs to be measured sensibly.

Summary

‘Collaborative filtering’

Ideas are broadly applicable. *Be creative!*

Clustering

K-means for minimizing ‘cluster variance’

Review notes, *not just slides*

[other methods exist: hierarchical, top-down and bottom-up]

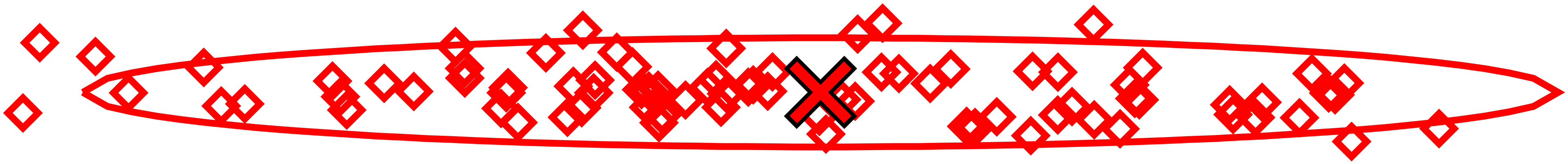
Unsupervised learning

Spot structure in unlabelled data

Combine with knowledge of task

Mixture modelling **(non-examinable)**

The fix: clusters have shapes as well as centers:



Assume each point is from one of K Gaussian distributions

Just like K-means, but:

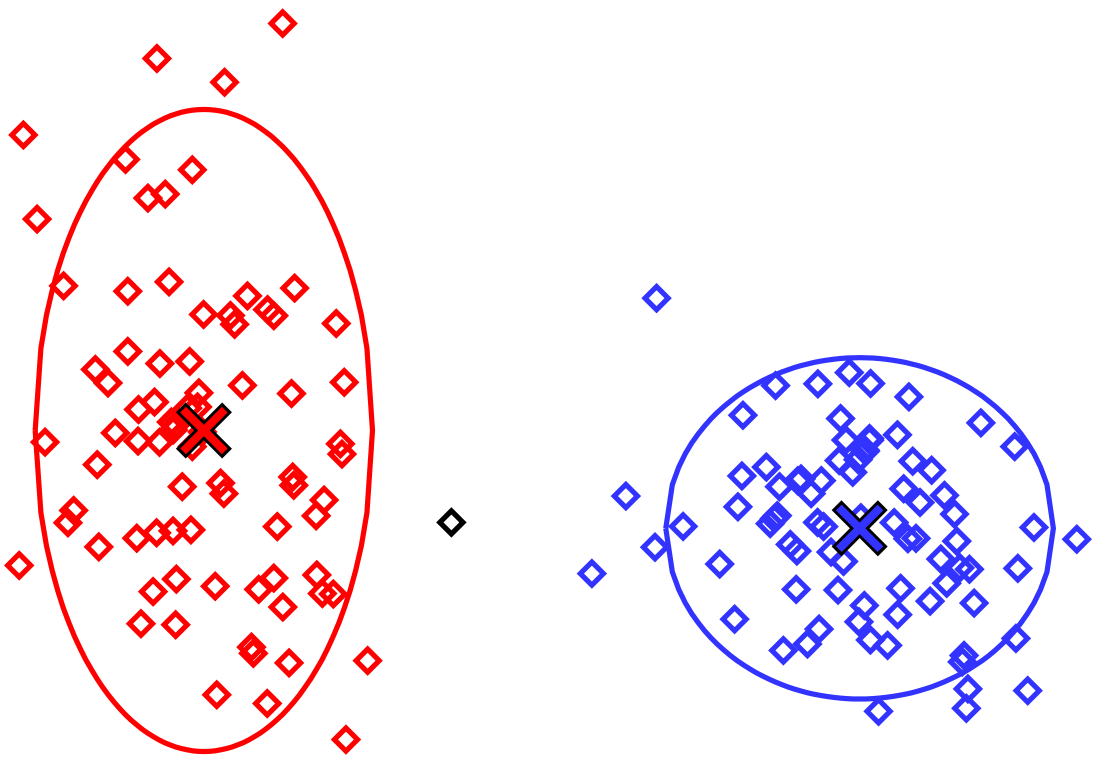
Assign points to Gaussian assigning highest probability.

Update cluster with mean and variance of points it owns.

Fancier (usual) version: points have soft assignments in proportion to their probability under each cluster.

Soft assignments

Each cluster $k \in \{1 \ldots K\}$ has fitted a model $P(\mathbf{x} \,|\, c=k)$.



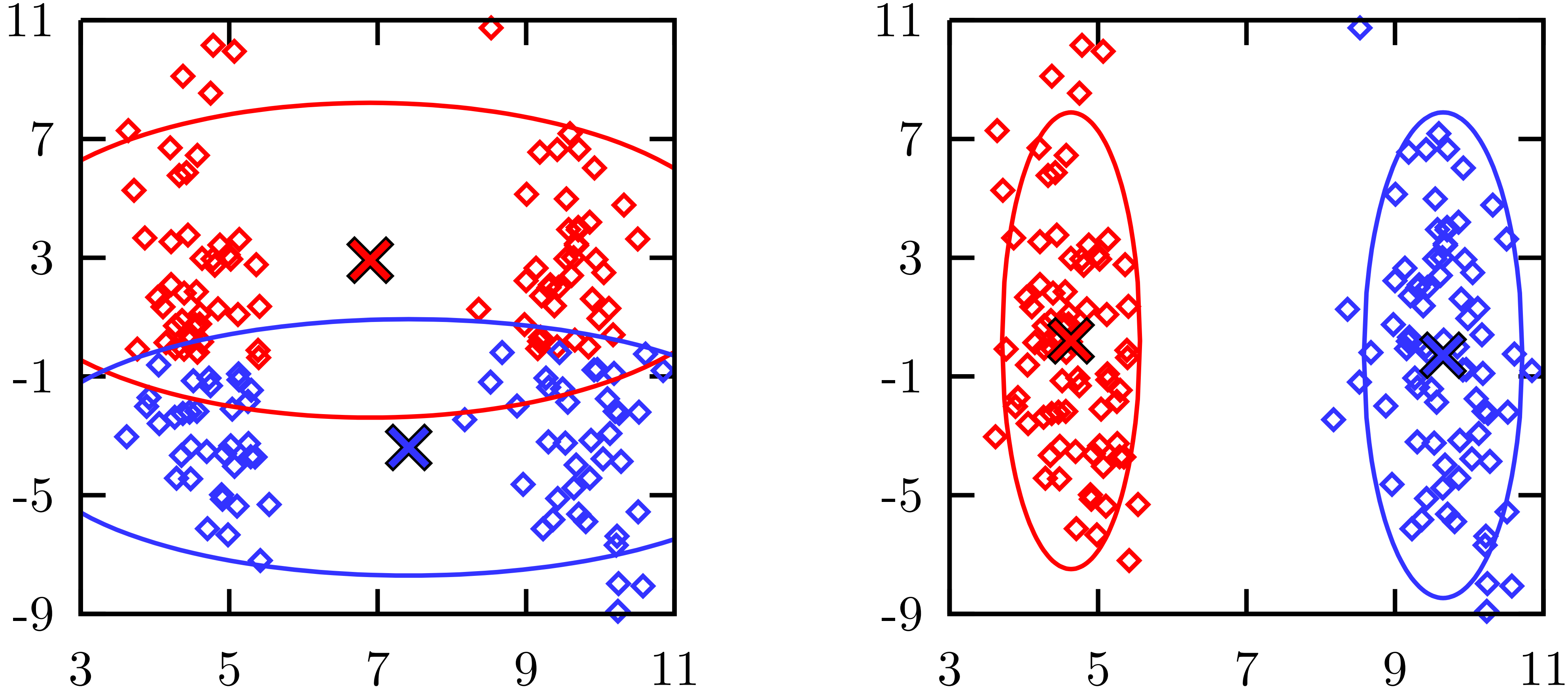
$$P(c=k \,|\, \mathbf{x}) = \frac{P(\mathbf{x} \,|\, c=k)\, P(c=k)}{P(\mathbf{x})} \propto P(\mathbf{x} \,|\, c=k)\, P(c=k)$$

Theory of mixture modelling

- **The model is called a mixture of Gaussians**
- **The algorithm is called EM** (Expectation Maximization) *
- EM maximizes $P(\text{data} \,|\, \text{fitted model})$
- Does EM converge?

* EM is a general method to maximize likelihoods of probabilistic models with *latent variables*, e.g. cluster assignments.

Fixing previous problems



The clustering on the right has much higher probability than the K-means solution on the left.