INTRODUCTION TO AI

LEGAL TECH HACKATHON 2023

TOINE BOGERS

IT UNIVERSITY OF CPH

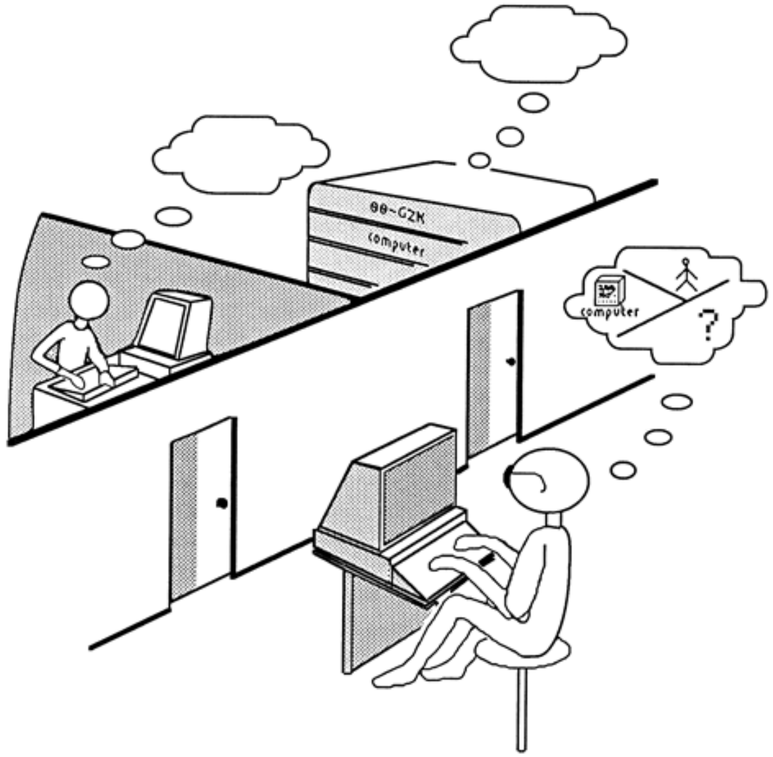
PIONEER CENTRE FOR ARTIFICIAL INTELLIGENCE

OUTLINE

- History of Artificial Intelligence
- Machine learning
- Solving problems with AI: CRISP-DM
- Example: Predicting person-job fit

HISTORY OF ARTIFICIAL INTELLIGENCE

HISTORY OF ARTIFICIAL INTELLIGENCE



Turing test (1950)

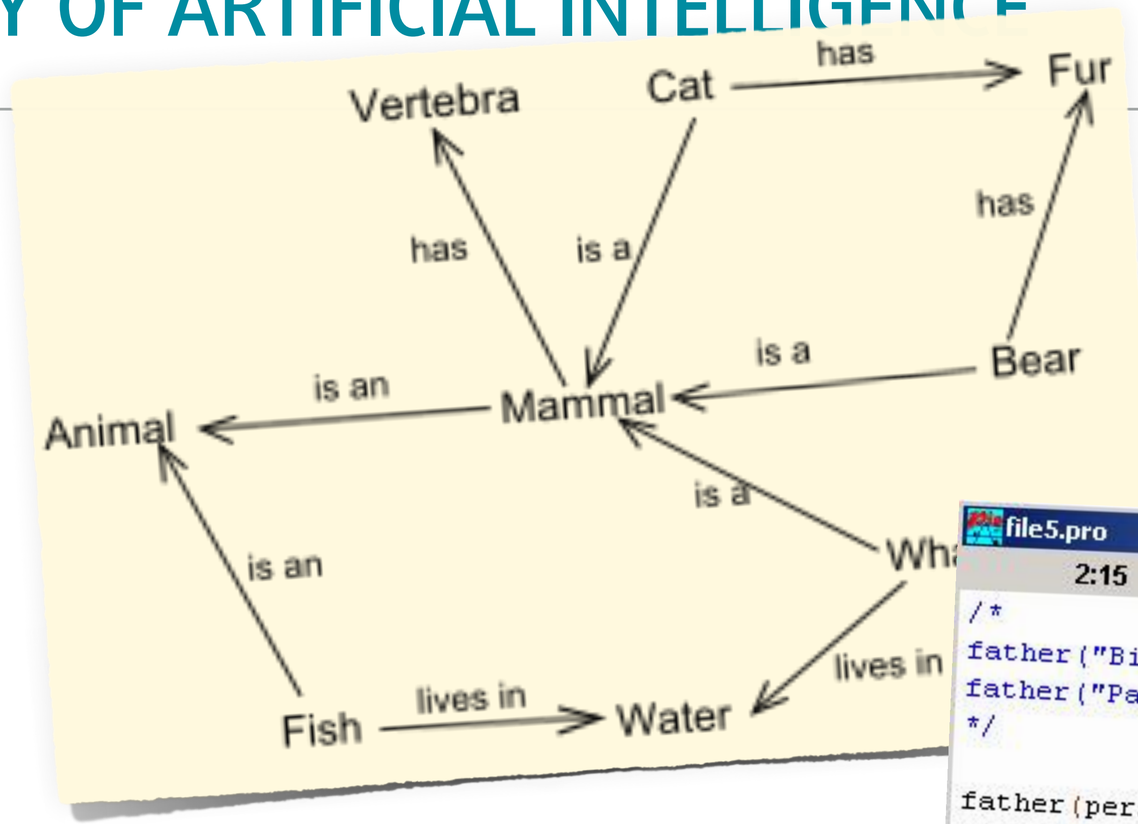
1950 1960 1970 1980 1990 2000 2010 2020 2030

HISTORY OF ARTIFICIAL INTELLIGENCE

Darthmouth AI summit (1956)

1950 1960 1970 1980 1990 2000 2010 2020 2030

HISTORY OF ARTIFICIAL INTELLIGENCE



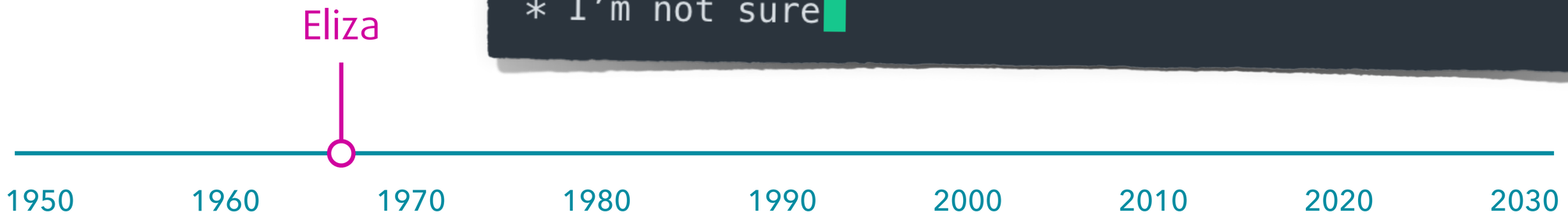

```
file5.pro
2:15    Insert    Indent
/*
father("Bill","John").
father("Pam","Bill").
*/

father(person("Bill","male"),person("John","male")).
father(person("Pam","female"),person("Bill","male")).

grandFather(Person,GrandFather):-
    father(Father,GrandFather),
    father(Person,Father).
```

Golden years (1956-1974)

1950 1960 1970 1980 1990 2000 2010 2020 2030

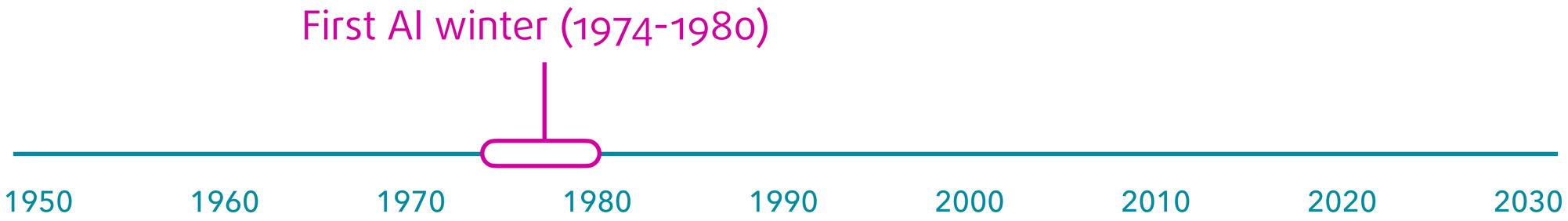
HISTORY OF ARTIFICIAL INTELLIGENCE


```
> Hello, I am Eliza.
* Hey!
> Don’t you ever say Hello?
* Hello!
> How are you today.. What would you like to discuss?
* I’m depressed.
> Do you enjoy being depressed?
* No, I don’t.
> Why no?
* Because I want to be happy.
> Why do you want to be happy?
* Because I’m tired od being depressed.
> That is quite interesting.
* I’m not sure
```



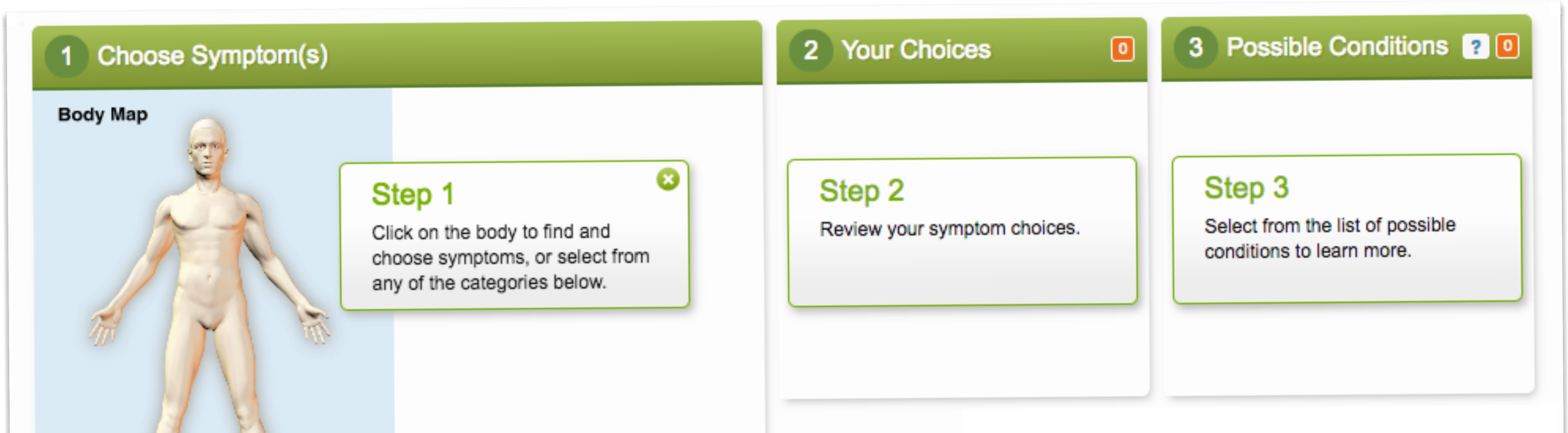
HISTORY OF ARTIFICIAL INTELLIGENCE

- Turns out AI was a difficult problem to solve!
  - Limited computing power
  - Importance of common-sense knowledge



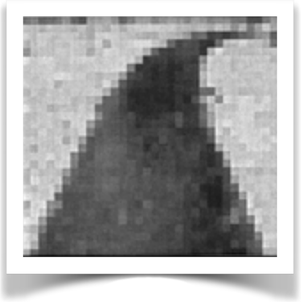
HISTORY OF ARTIFICIAL INTELLIGENCE

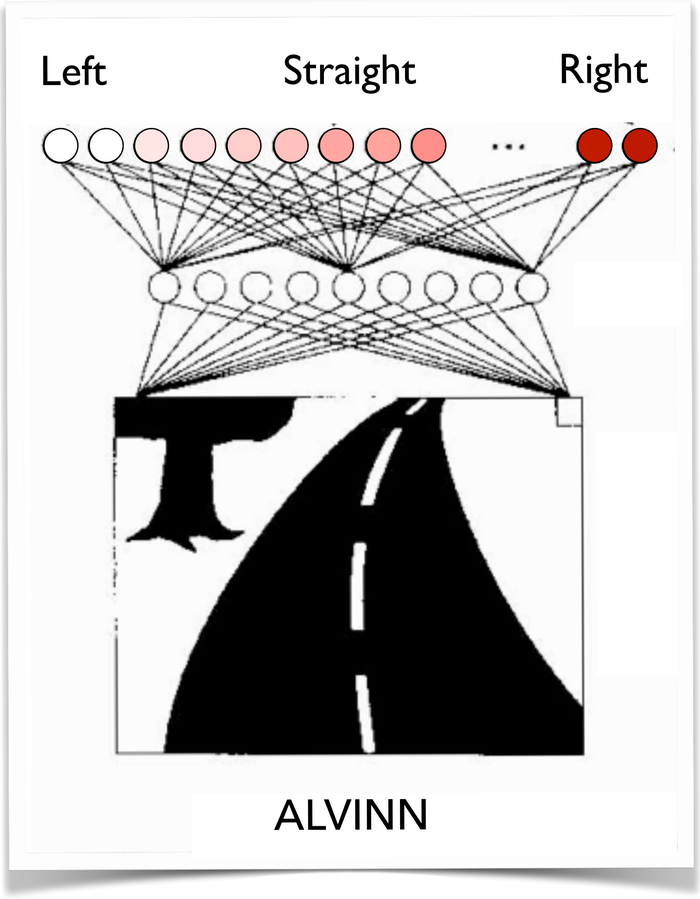
- Second wave of funding for AI projects
  - **Expert systems** (= top-down AI)
  - Collections of handcrafted rules representing expert intelligence
  - Example: WebMD Symptom Checker (http://symptoms.webmd.com/)



HISTORY OF ARTIFICIAL INTELLIGENCE

Input: 32 x 32 pixels



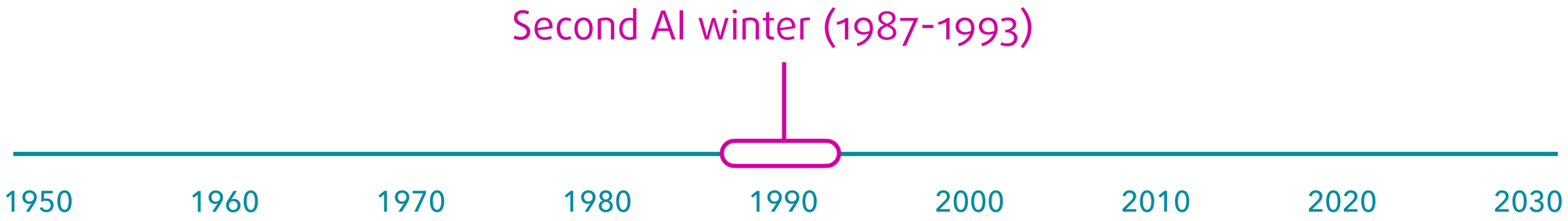


ALVINN (1986)

1950 1960 1970 1980 1990 2000 2010 2020 2030

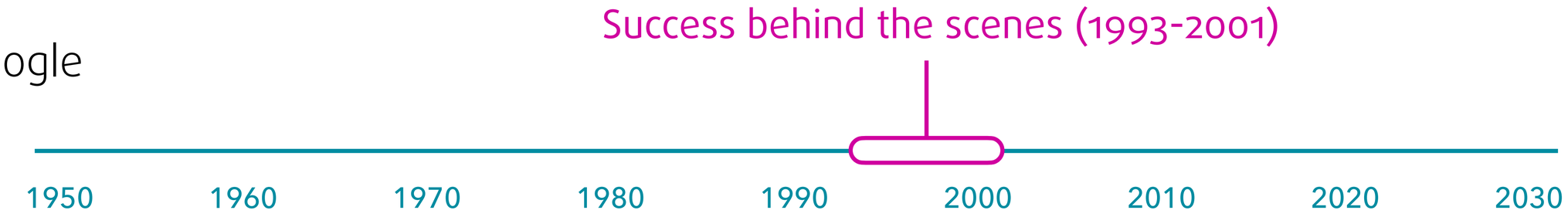
HISTORY OF ARTIFICIAL INTELLIGENCE

- Top-down AI did not fulfil its promises
  - **Tacit** knowledge is hard to encode
  - Maintenance costs
  - Inflexible



HISTORY OF ARTIFICIAL INTELLIGENCE

- Behind-the-scenes success of AI
  - Speech recognition
  - Data mining
  - Industrial robotics
  - Banking software
  - Medical diagnosis
  - Google



HISTORY OF ARTIFICIAL INTELLIGENCE

- New golden age for AI research
  - Massive amounts of data available
  - Computing power to cope with it
  - New breakthroughs in artificial neural networks (= **deep learning**)

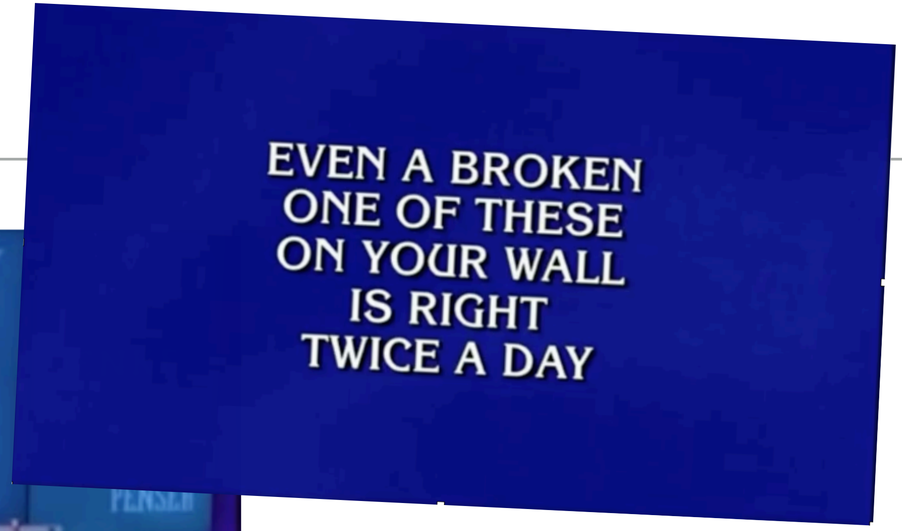
HISTORY OF ARTIFICIAL INTELLIGENCE

Autonomous aerial helicopter maneuvers (2008)

DARPA's self-driving car challenge (2004)

1950 1960 1970 1980 1990 2000 2010 2020 2030

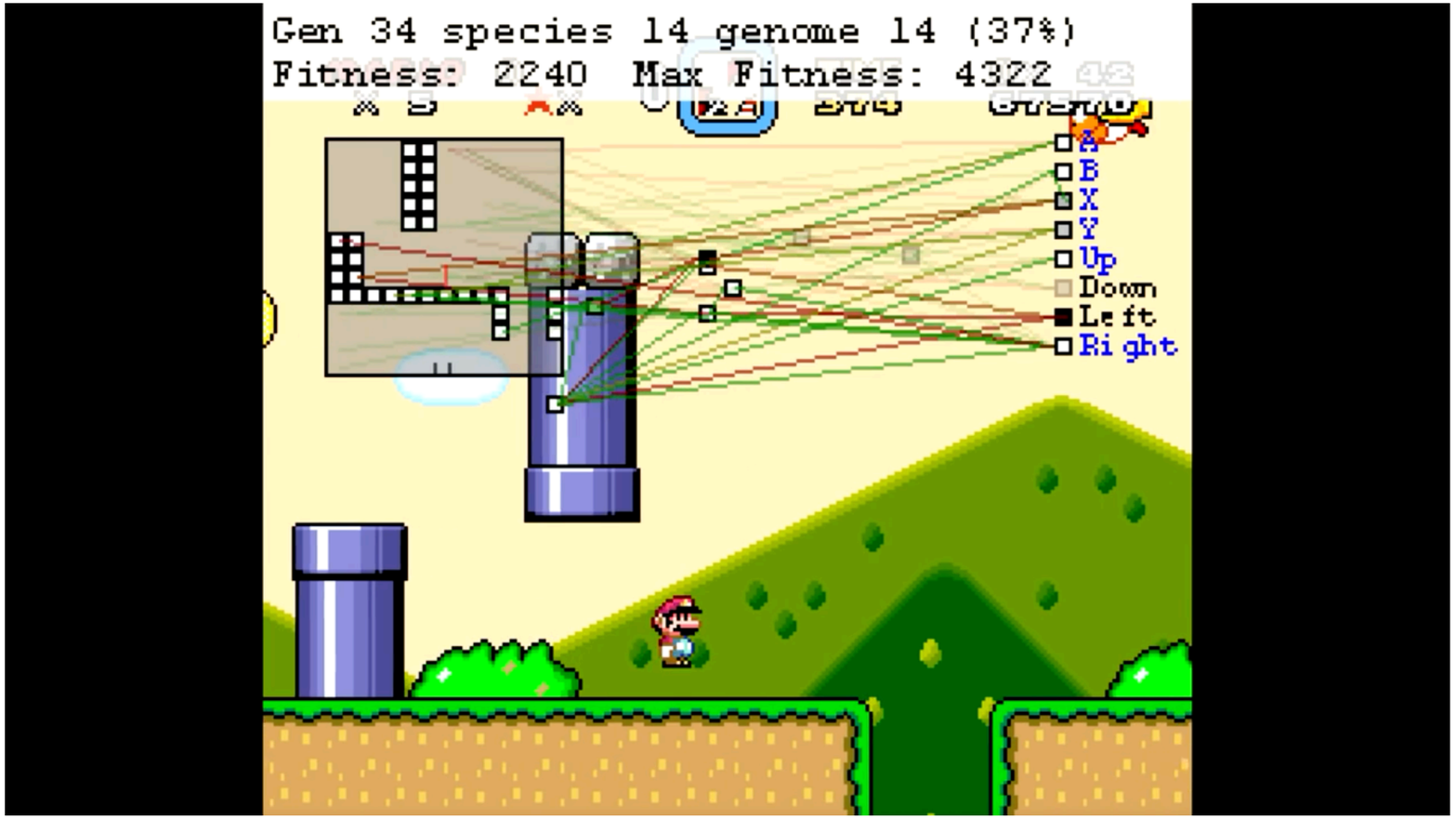
HISTORY OF ARTIFICIAL INTELLIGENCE



IBM Watson (2011)

1950 1960 1970 1980 1990 2000 2010 2020 2030

HISTORY OF ARTIFICIAL INTELLIGENCE



MarI/O (2015)

1950 1960 1970 1980 1990 2000 2010 2020 2030

HISTORY OF ARTIFICIAL INTELLIGENCE

- Feels like we are entering a new era: **generative AI**
  - Research on generative AI has been going on for years (or decades even), but ChatGPT and DALL-E have catapulted this into the limelight
  - Huge potential **impact on future of work**
  - Also an era with an increased focus on **responsible** use of AI
    - Fairness, transparency, trustworthiness, accountability, ethics, …

HISTORY OF ARTIFICIAL INTELLIGENCE

- Want to know more? Watch the Royal Institute's documentary on the past, present & future of AI
  - https://youtu.be/8FHBh_OmdsM

MACHINE LEARNING

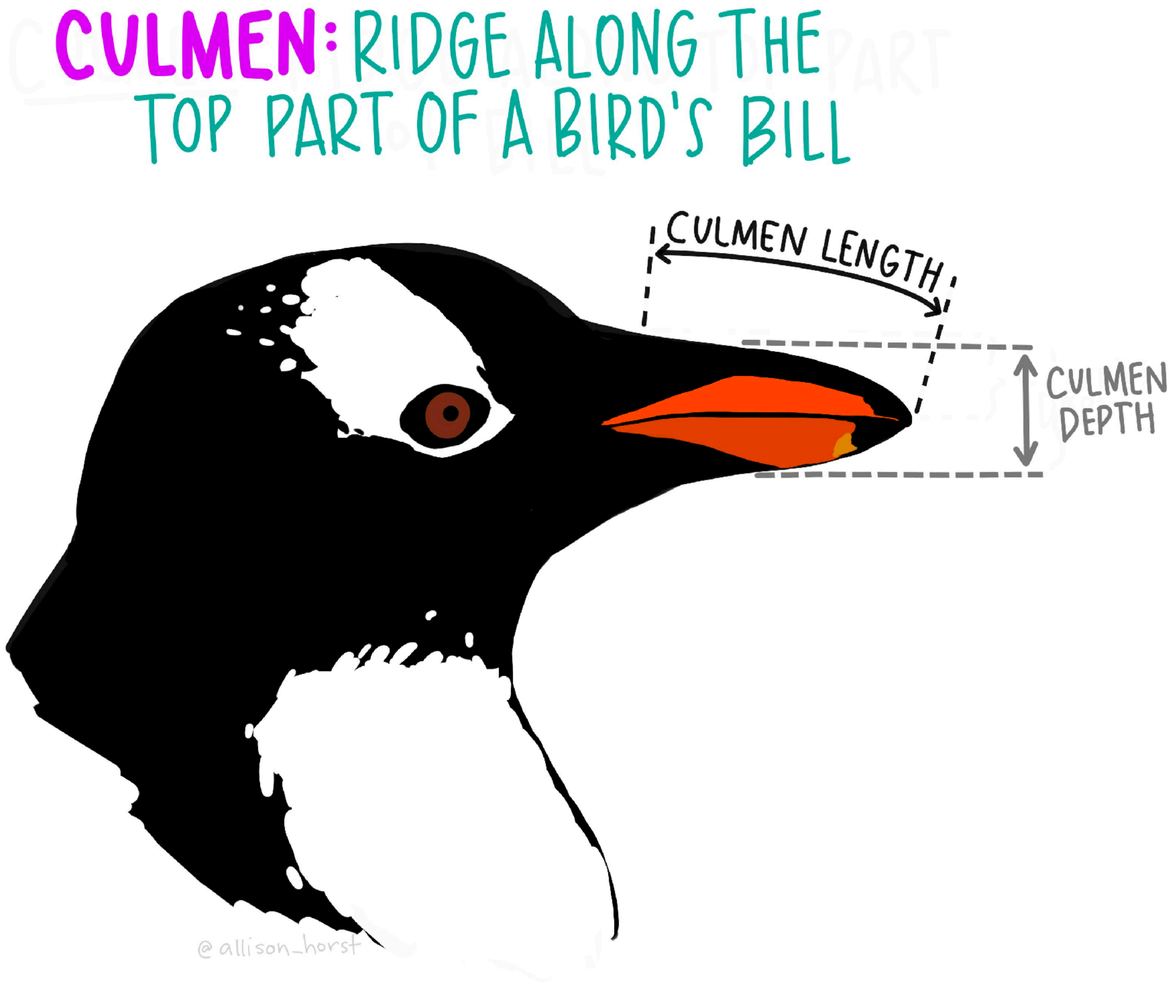
APPROACHES TO ARTIFICIAL INTELLIGENCE

- Artificial intelligence involves
  - Identifying **relevant aspects & patterns** of the problem
  - **Learning how they are related** to real-world outcomes
- Historical dichotomy of AI approaches
  - **Top-down** approaches
    - We tell the computer exactly how it should solve a task (e.g., expert systems)
    - Example: `IF <SUBJECT> contains ‘viagra’ THEN mark as SPAM`
  - **Bottom-up** approaches
    - The computer finds out for itself with a ‘little’ help from us (e.g., **machine learning**)
    - Example: The word ‘`viagra`’ is very often found in spam e-mails

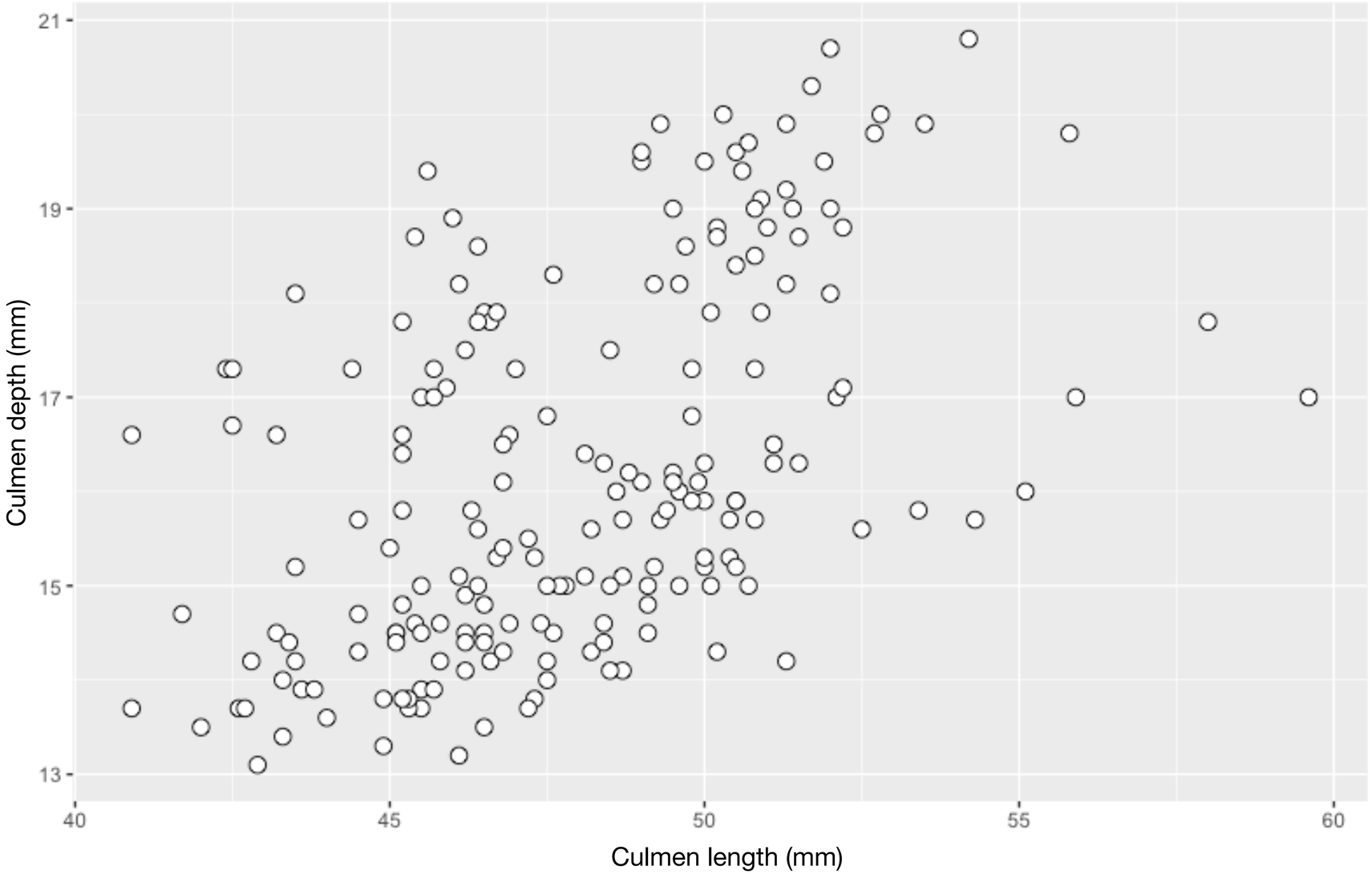
APPROACHES TO ARTIFICIAL INTELLIGENCE

- Two important types of machine learning approaches
  - **Supervised learning**
    - Problems where a **target variable** exists that we wish to **predict** or **classify** data into
    - Computer **learns from examples** with **correct answers** that we provide
    - Learn a model that identifies commons patterns in the input data and maps those to an output
      - ★ When output is a category (= discrete variable) → **classification**
      - ★ When output is a continuous variable → **regression**
  - **Unsupervised learning**
    - Goal of the computer is to identify common patterns to **cluster** together similar data points and detect the most important categories
    - In other words: trying to classify the data without knowing the answers beforehand

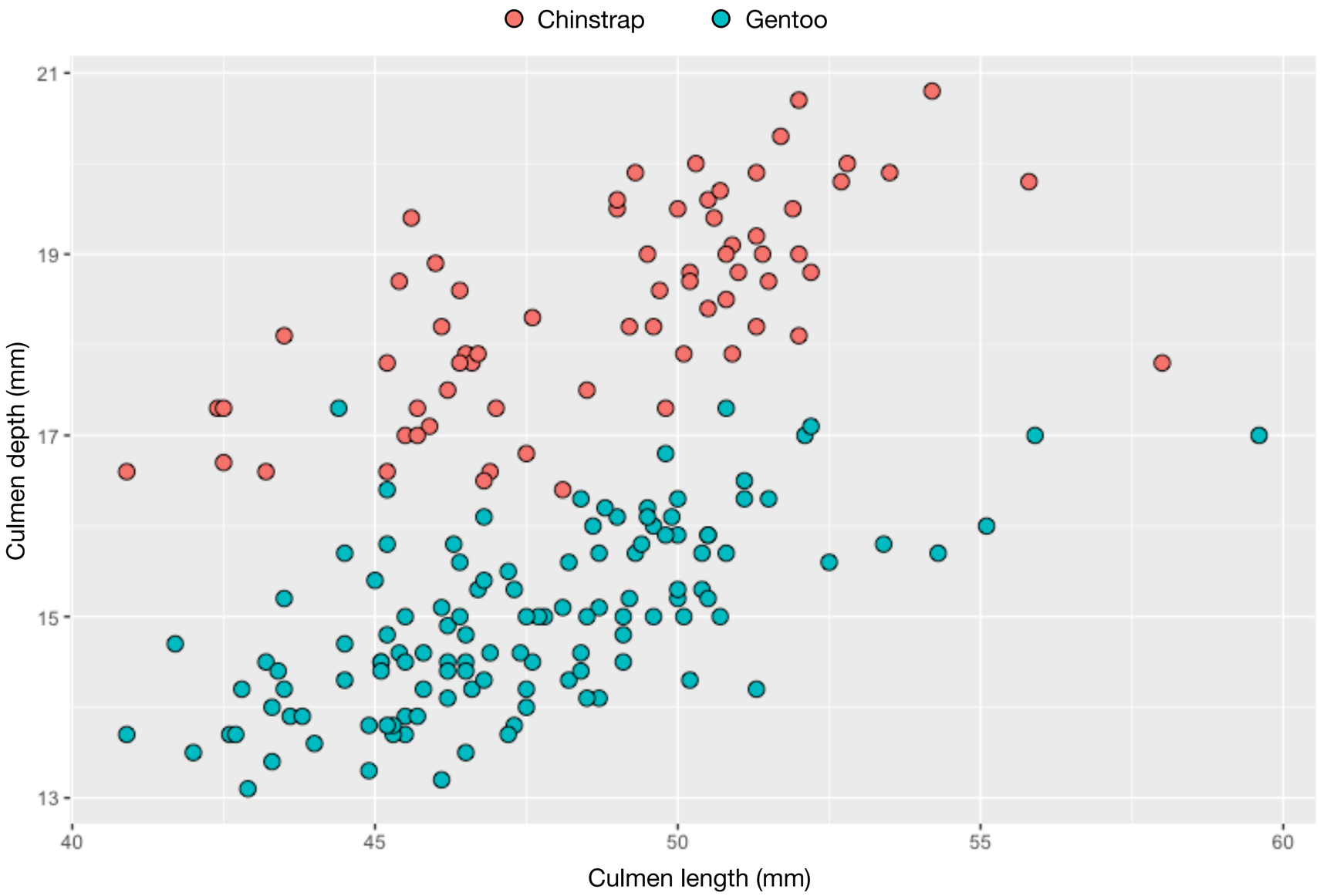
EXAMPLE: PALMER'S PENGUINS



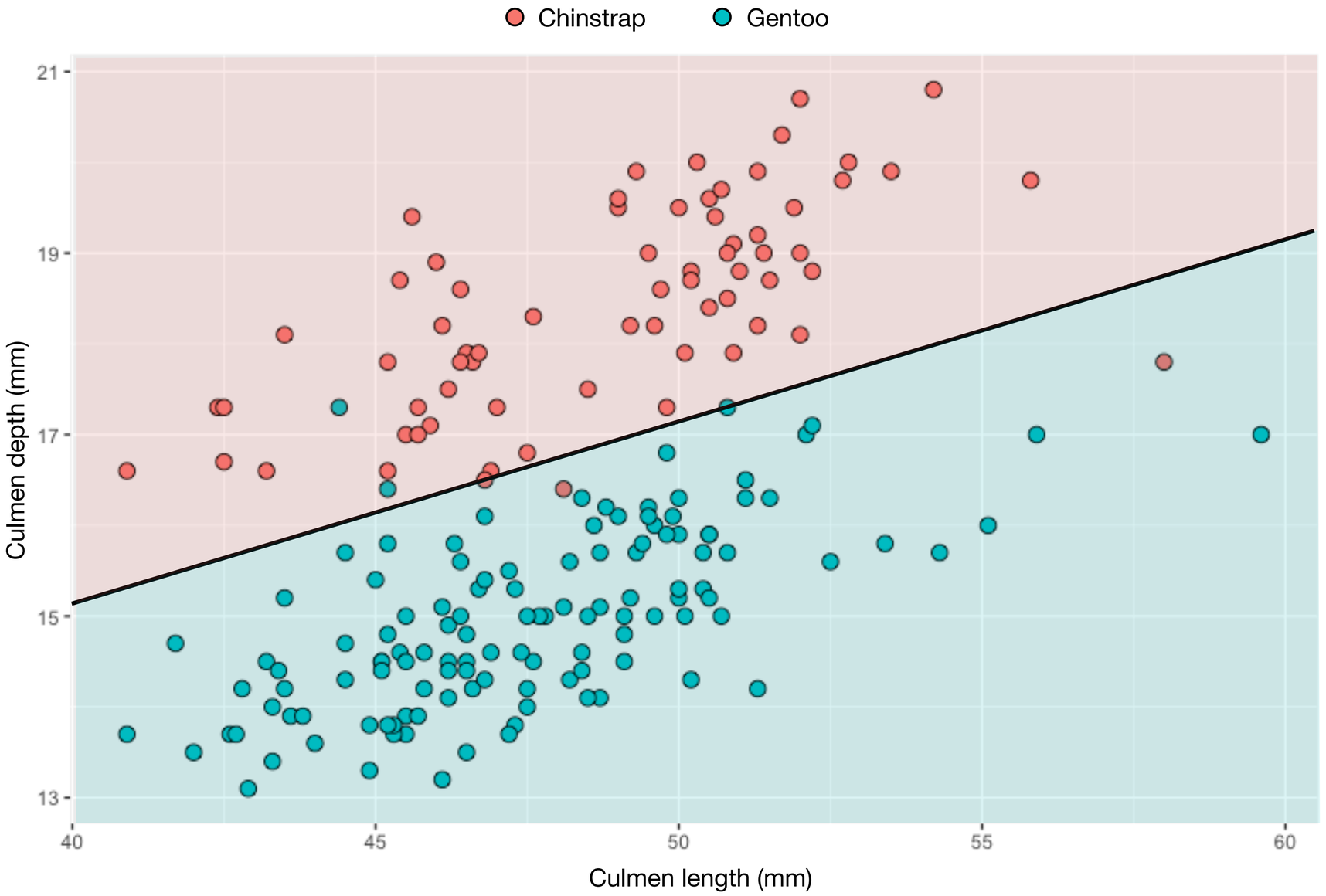
EXAMPLE: PALMER'S PENGUINS



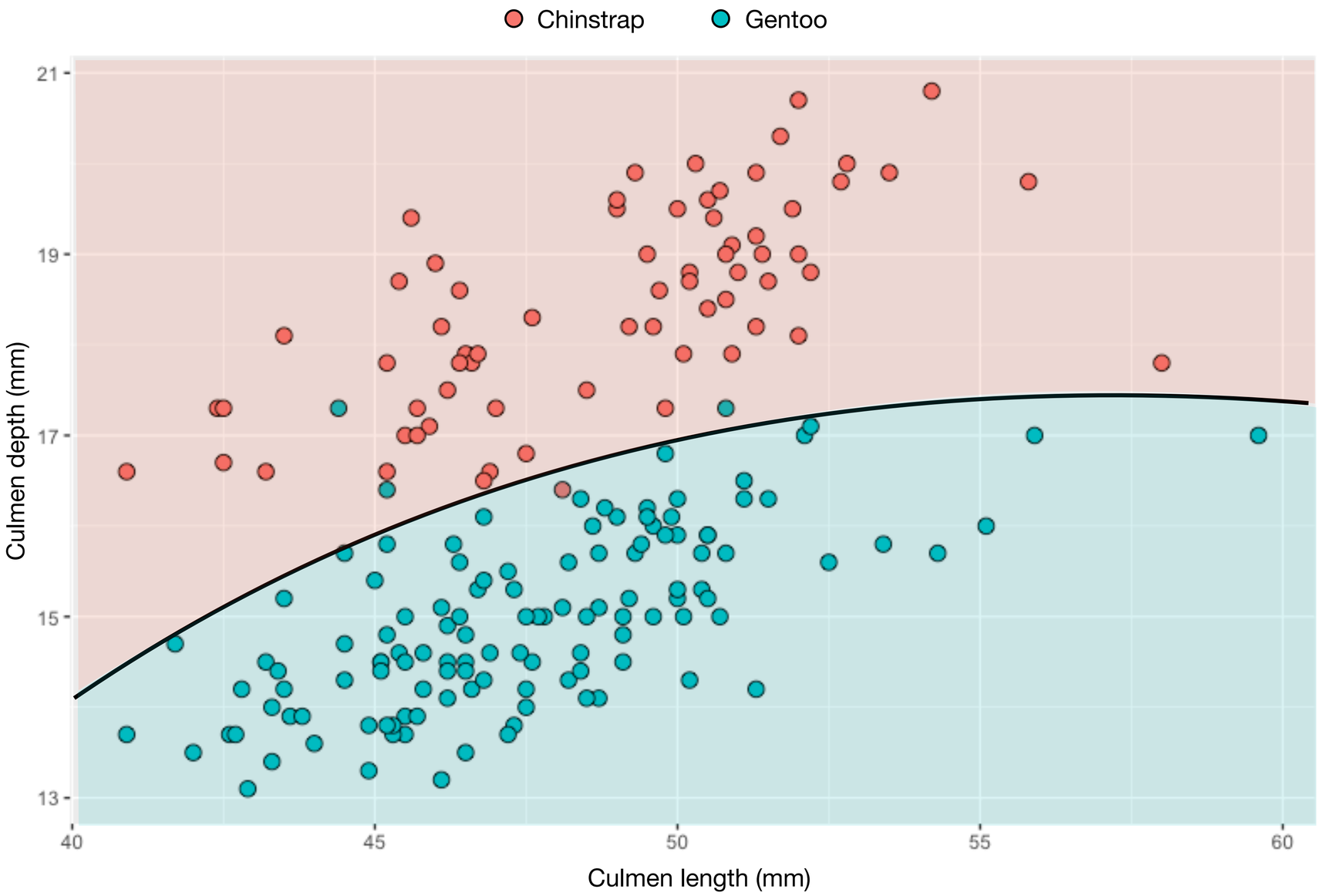
CLASSIFICATION IS ABOUT DIVIDING UP OUR DATA



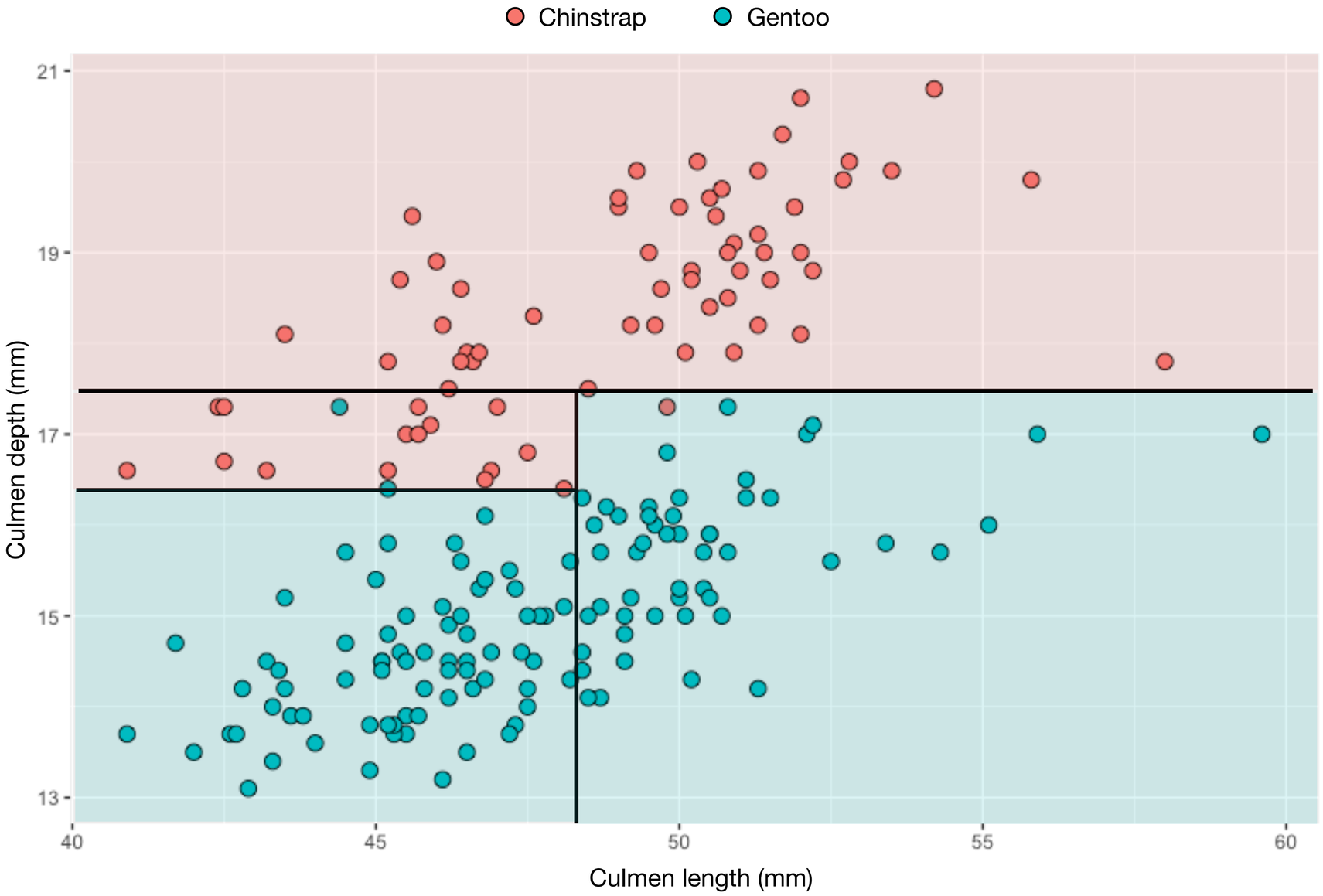
LINEAR


Chinstrap
Gentoo
21
19
17
15
13
Culmen depth (mm)
40
45
50
55
60
Culmen length (mm)

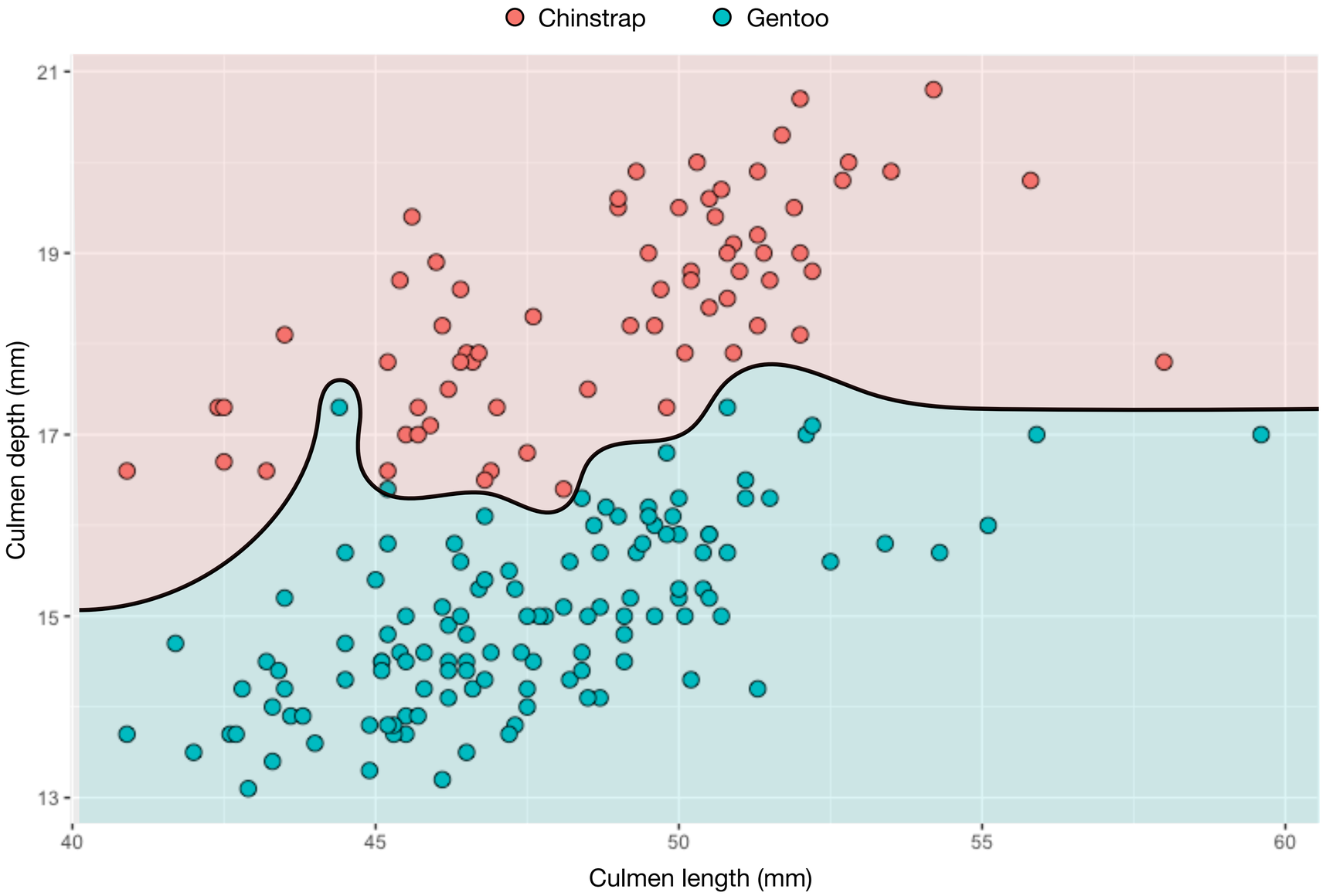
POLYNOMIAL


Chinstrap
Gentoo
21
19
17
15
13
Culmen depth (mm)
40
45
50
55
60
Culmen length (mm)

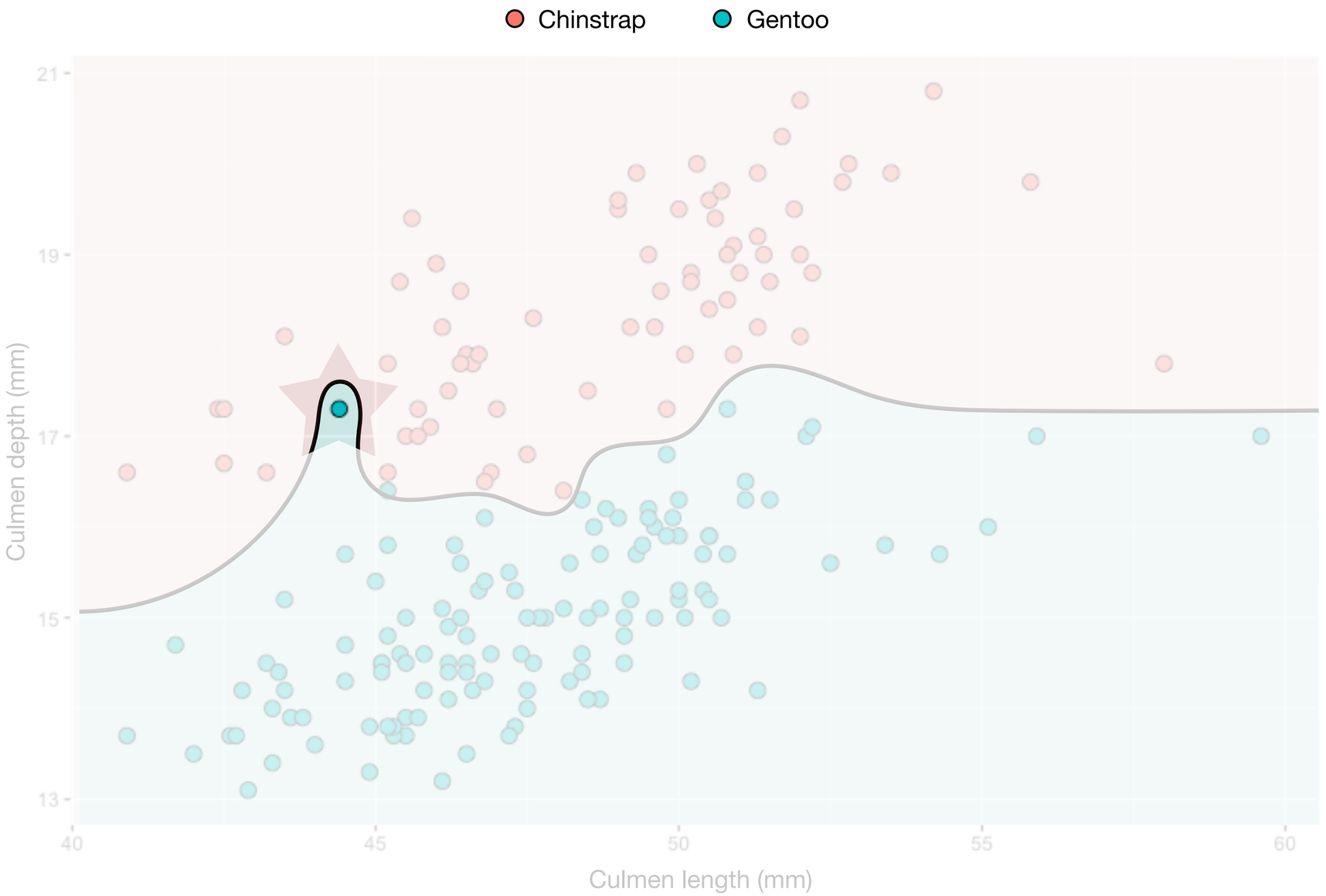
DECISION TREES



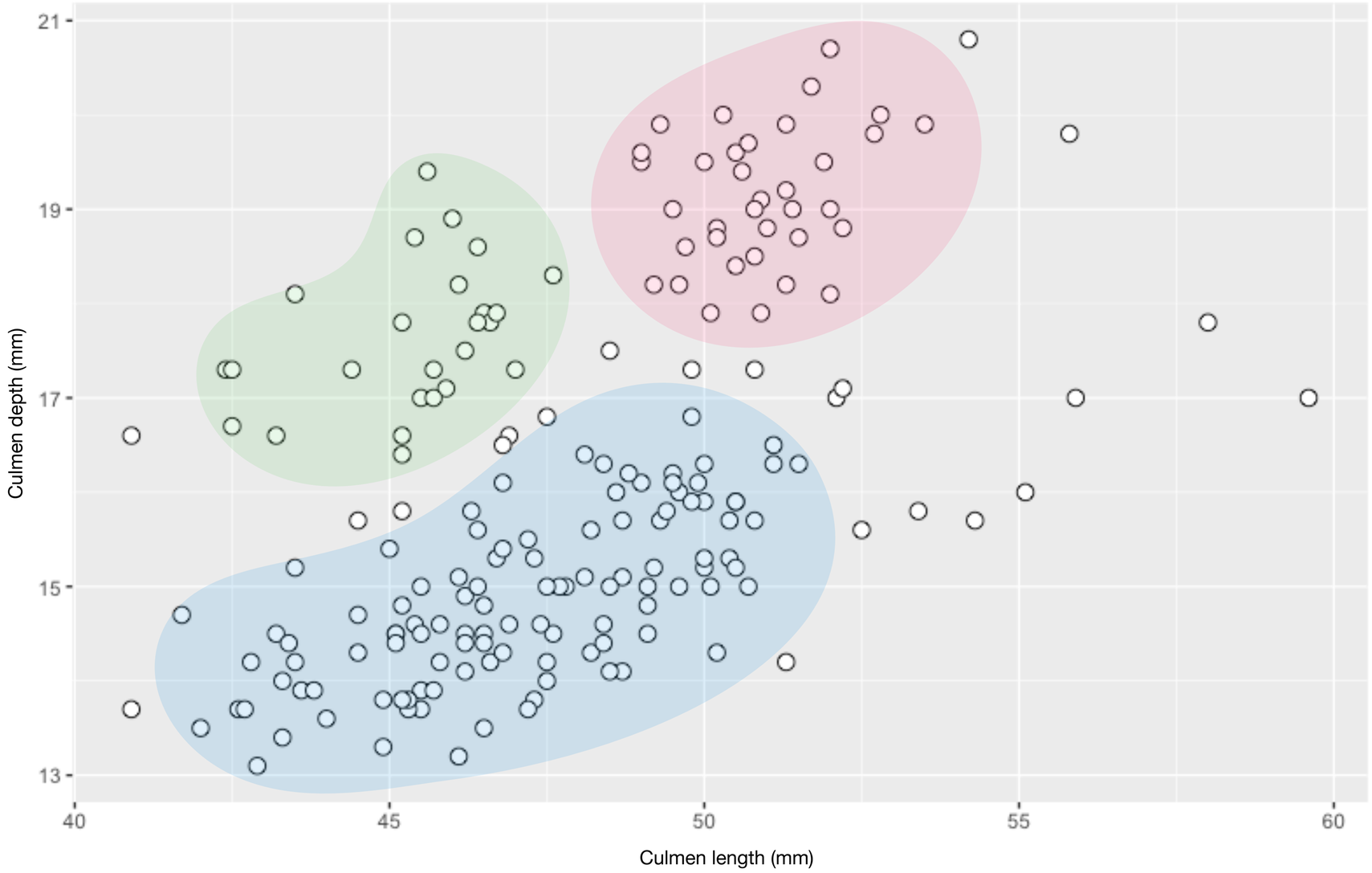
DEEP LEARNING



WHAT ABOUT THIS PENGUIN?



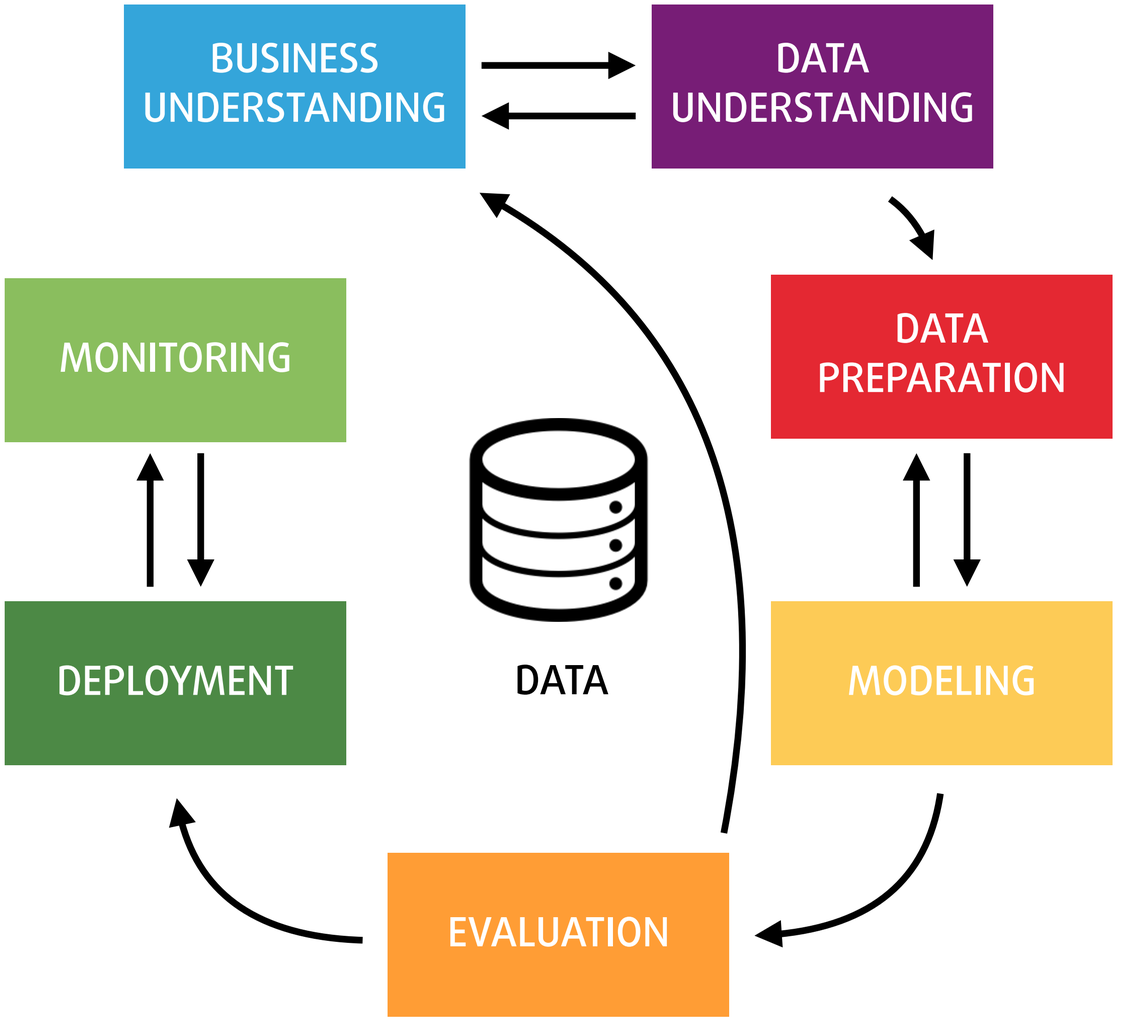
CLUSTERING IS ABOUT FINDING THE GROUPS



SOLVING PROBLEMS WITH AI: CRISP-DM

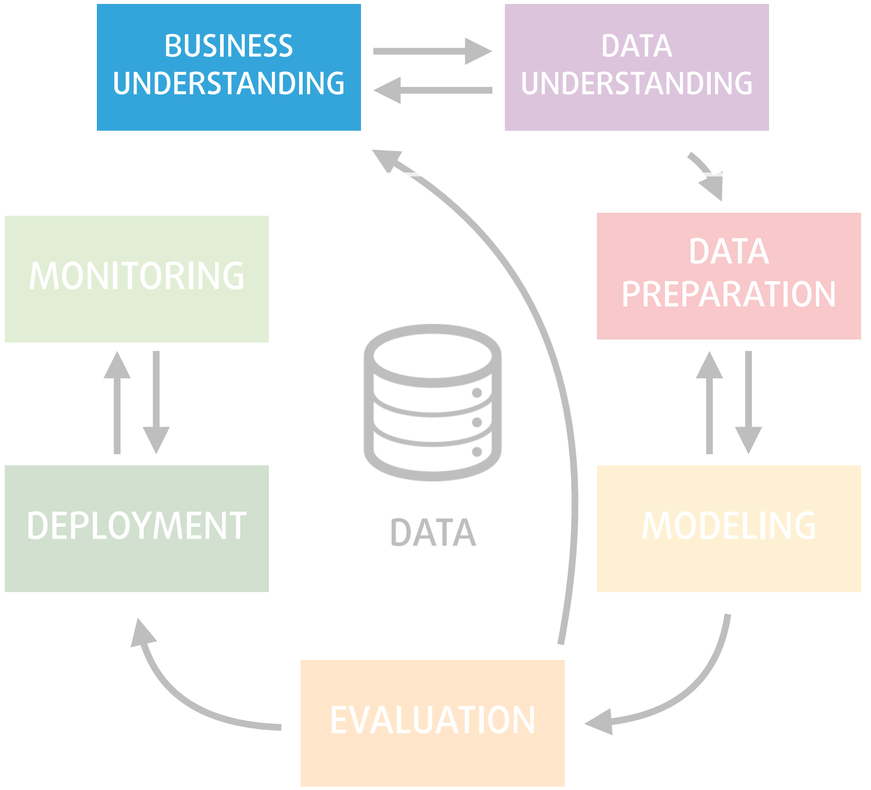
CRISP-DM

- **CRISP-DM** (Cross Industry Standard Process for Data Mining) is a process model that describes commonly-used approaches to solving problems with machine learning
  - Business understanding
  - Data understanding
  - Data preparation
  - Modeling
  - Evaluation
  - Deployment
  - Monitoring



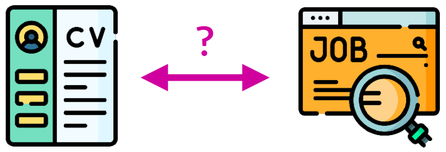
BUSINESS UNDERSTANDING

- **Business understanding**
  - Understanding the **project objectives and requirements** from a business perspective
  - Converting this knowledge into a **preliminary plan** designed to achieve the objectives
- **Example scenario:** matching jobs and job seekers
  - We want to use AI to **match job vacancies to job seekers**
  - Benefits: saving recruiters' time, contacting more relevant candidates, better matches

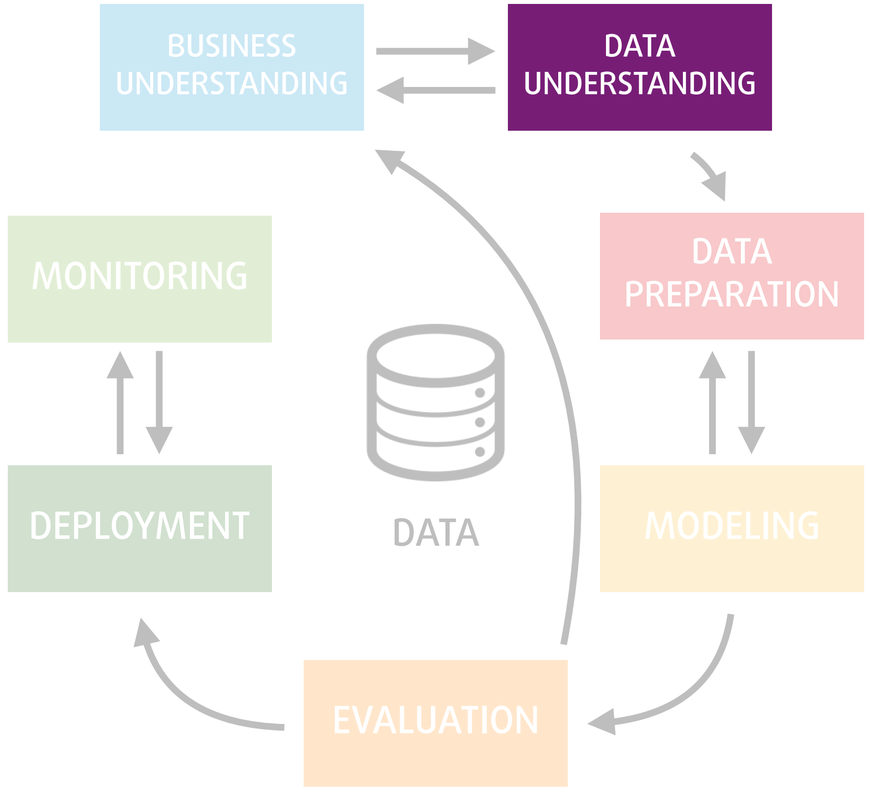


DATA UNDERSTANDING

- **Data understanding**
  - **What** are we trying to predict?
    - Is a job seeker a match for the job?

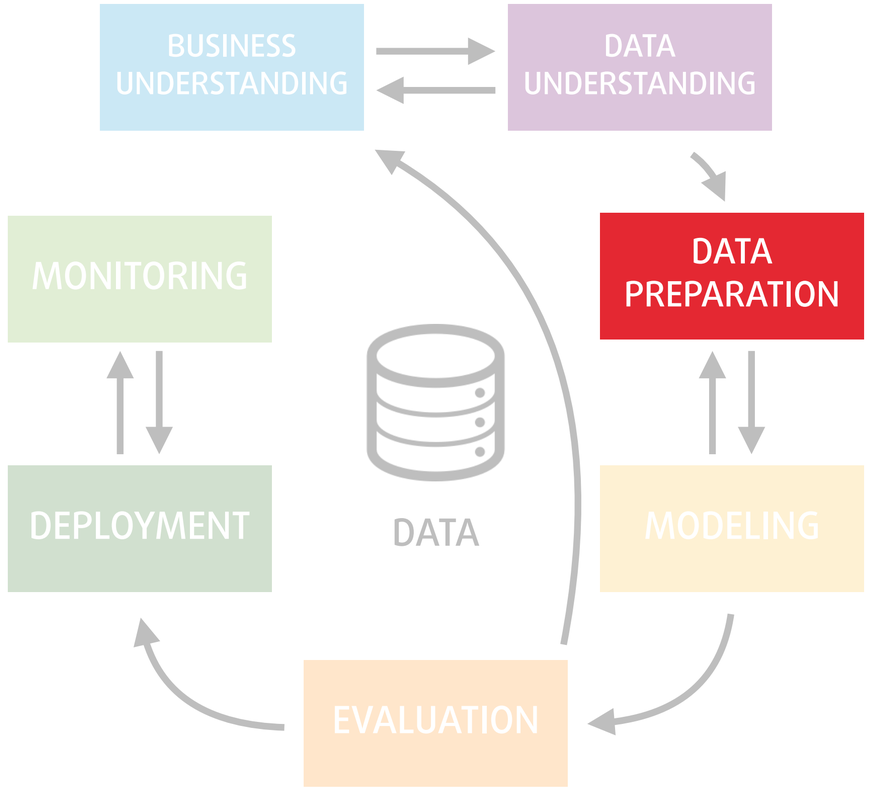


  - **What kind of data** do we need for this?
    - Skills & competences, job titles, work experience, management experience, education, salary, part-time vs. full-time
    - Examples of successful (and unsuccessful) matches
  - **Where** could we find such data to **train** our prediction system?
    - Past CVs and job ads
    - Past selections by recruiters
    - Past responses by job seekers



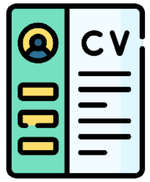
DATA PREPARATION

- **Data preparation** covers all activities to construct the final data set from the initial raw data
  - Data sampling/selection
  - Data cleaning
  - Data transformation
  - Feature construction
    - Extracting skills & competences from the job ads and CVs
    - Extracting part-time vs. full-time status
  - Outlier detection & removal



DATA PREPARATION

Job title	Required skills	Min. years of work experience
Veterinary assistant	Communication, empathy, time management, experience machine X	3



Past job titles	Skills	Years of work experience	Relevant education?
Veterinary assistant, volunteer at animal hospital	Team player, passion for animals, time management, experience with machine Y	5	Yes

From recruiter selections or job seeker feedback

Match?
Yes

TRAINING EXAMPLE

Job title	Required skills	Min. years of work exp.	Past job titles	Skills	Years of work experience	Relevant education?	Match?
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant, volunteer at animal hospital	Team player, passion for animals, time management, experience with	5	Yes	Yes

FEATURES

Job title	Required skills	Min. years of work exp.	Past job titles	Skills	Years of work experience	Relevant education?	Match?
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant, volunteer at animal hospital	Team player, passion for animals, time management, experience with	5	Yes	Yes

TRAINING SET

Job title	Required skills	Min. years of work exp.	Past job titles	Skills	Years of work experience	Relevant education?	Match?
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant, volunteer at animal hospital	Team player, passion for animals, time management, experience with	5	Yes	Yes
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Volunteer dog walker, shop assistant	Passion for animals, customer-oriented	0	No	No

TRAINING SET

Job title	Required skills	Min. years of work exp.	Past job titles	Skills	Years of work experience	Relevant education?	Match?
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant, volunteer at animal hospital	Team player, passion for animals, time management, experience with	5	Yes	Yes
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Volunteer dog walker, shop assistant	Passion for animals, customer-oriented	0	No	No
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant	Experience with machine X, passion for animals	14	Yes	Yes

TRAINING SET

Job title	Required skills	Min. years of work exp.	Past job titles	Skills	Years of work experience	Relevant education?	Match?
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant, volunteer at animal hospital	Team player, passion for animals, time management, experience with	5	Yes	Yes
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Volunteer dog walker, shop assistant	Passion for animals, customer-oriented	0	No	No
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant	Experience with machine X, passion for animals	14	Yes	Yes
Tech support	Fluent in English, patient, iOS, Android, mobile phones	2	IT support staff	iOS, macOS, Windows	4	No	Yes

TEST CASE

Job title	Required skills	Min. years of work exp.	Past job titles	Skills	Years of work experience	Relevant education?	Match?
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant, volunteer at animal hospital	Team player, passion for animals, time management, experience with	5	Yes	Yes
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Volunteer dog walker, shop assistant	Passion for animals, customer-oriented	0	No	No
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant	Experience with machine X, passion for animals	14	Yes	Yes
Tech support	Fluent in English, patient, iOS, Android, mobile phones	2	IT support staff	iOS, macOS, Windows	4	No	Yes
Tech support	Fluent in English, patient, iOS, Android, mobile phones	2	Head of IT support	iOS, macOS, Windows, team management, networking, remote deployment	13	Yes	?

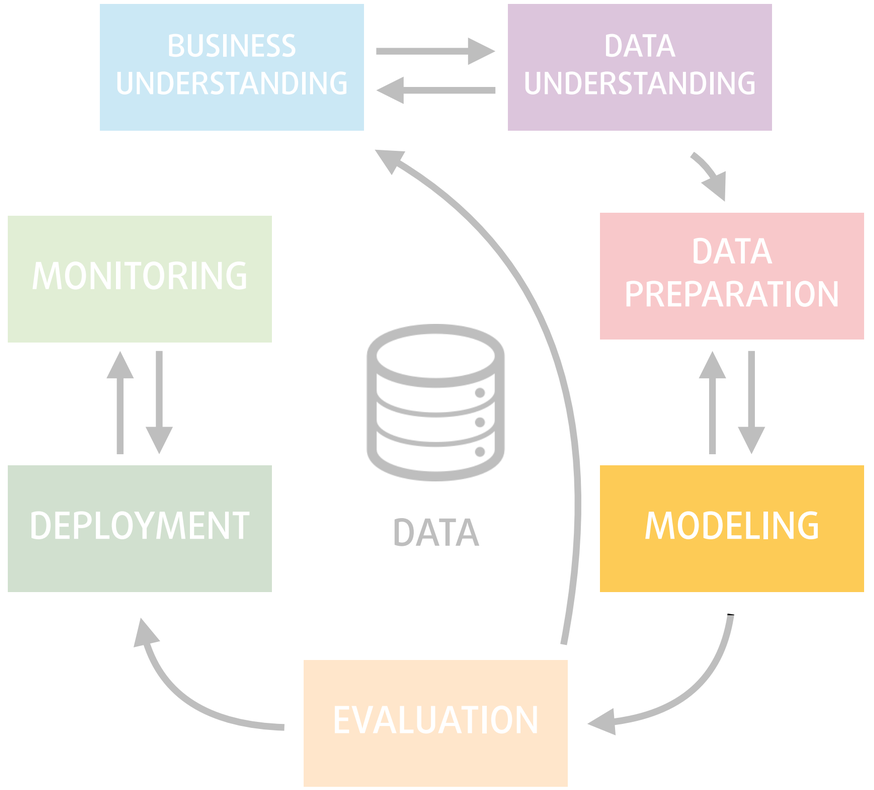
TEST CASE

Job title	Required skills	Min. years of work exp.	Past job titles	Skills	Years of work experience	Relevant education?	Match?
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant, volunteer at animal hospital	Team player, passion for animals, time management, experience with	5	Yes	Yes
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Volunteer dog walker, shop assistant	Passion for animals, customer-oriented	0	No	No
Veterinary assistant	Communication, empathy, time management, experience machine X	3	Veterinary assistant	Experience with machine X, passion for animals	14	Yes	Yes
Tech support	Fluent in English, patient, iOS, Android, mobile phones	2	IT support staff	iOS, macOS, Windows	4	No	Yes

Tech support	Fluent in English, patient, iOS, Android, mobile phones	2	Head of IT support	iOS, macOS, Windows, team management, networking, remote deployment	13	Yes	**Prediction: No**

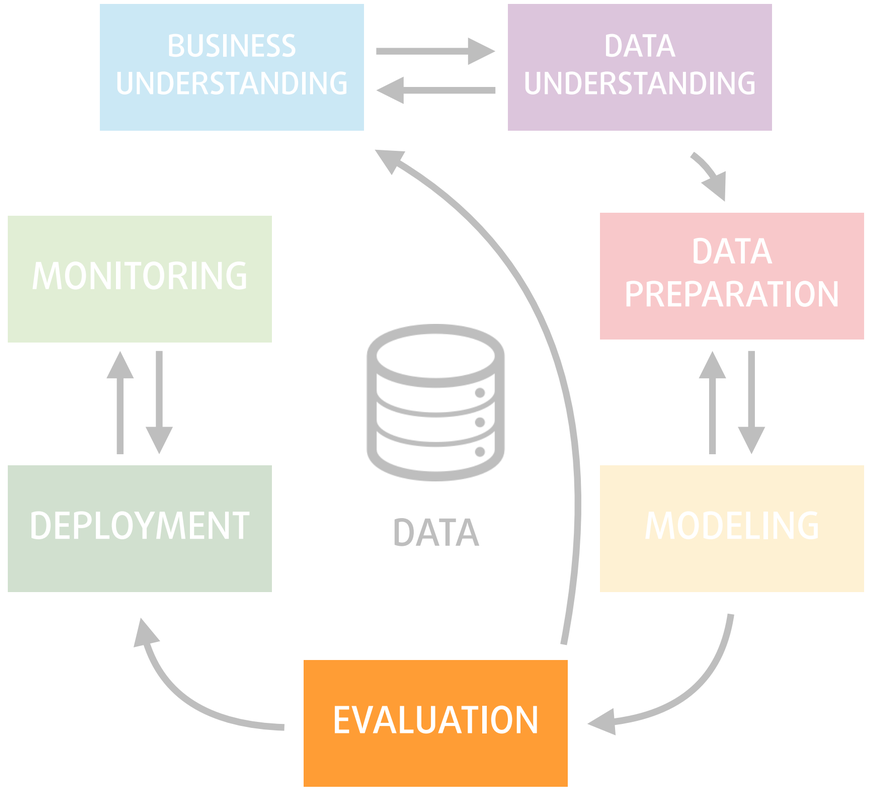
MODELING

- **Modeling**
  - Learn a model that identifies **commons patterns** in the **training** data
  - Map those patterns to an **output** (match or no match)
  - Many different machine learning models exist, each with their own strengths and weaknesses



EVALUATION

- Model(s) need to be **evaluated** before final deployment
  - How well does it **achieve business objectives**?
  - How well does it **predict the target classes**?
    - How **often** do we make mistakes?
    - How do we **use our training data** to evaluate this?
    - Are all mistakes equally bad?
  - We evaluate our model by performing so-called **backtesting**
    - We hide the actual answers in the training set from our classifier
    - We then compare the predictions to the actual answers
    - The more the algorithm gets right, the better!



MAIN TAKEAWAYS

- When attacking your case, consider the following
  - What is the (business) problem you are solving?
  - What data do you need to solve this problem?
  - Where can you get this data?
  - What needs to be done to make this data usable (e.g., cleaning, sampling, filtering)?
  - How will you know whether you are doing a good job?

QUESTIONS?