

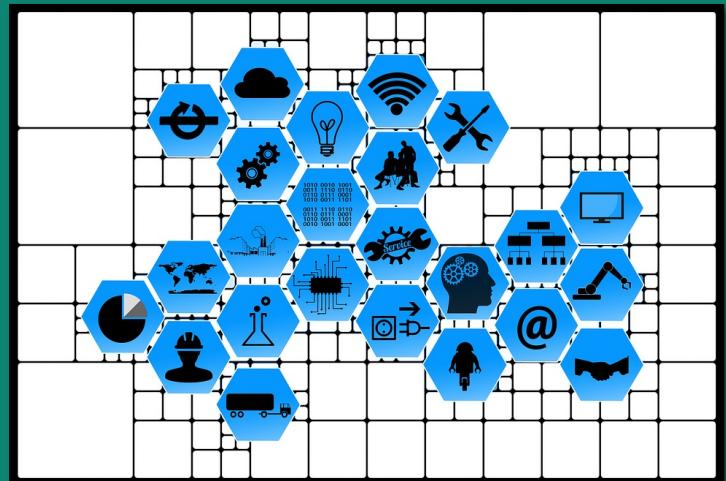


SIT123

Data Capture Technologies

Dr Imali Dias
imali.dias@deakin.edu.au

Lecture –Week 3
Trimester 2, 2022



Introduction to SIT123 Data Capture Technologies

- Saving data
- Analyzing data



DEAKIN
UNIVERSITY

Recap: what we have done so far

Week 1:

Sense-Think-Act

Arduino Blink – setup() &
loop()

Week 2:

Sensors

Sampling rate, accuracy, sensitivity

Analog & Digital signals

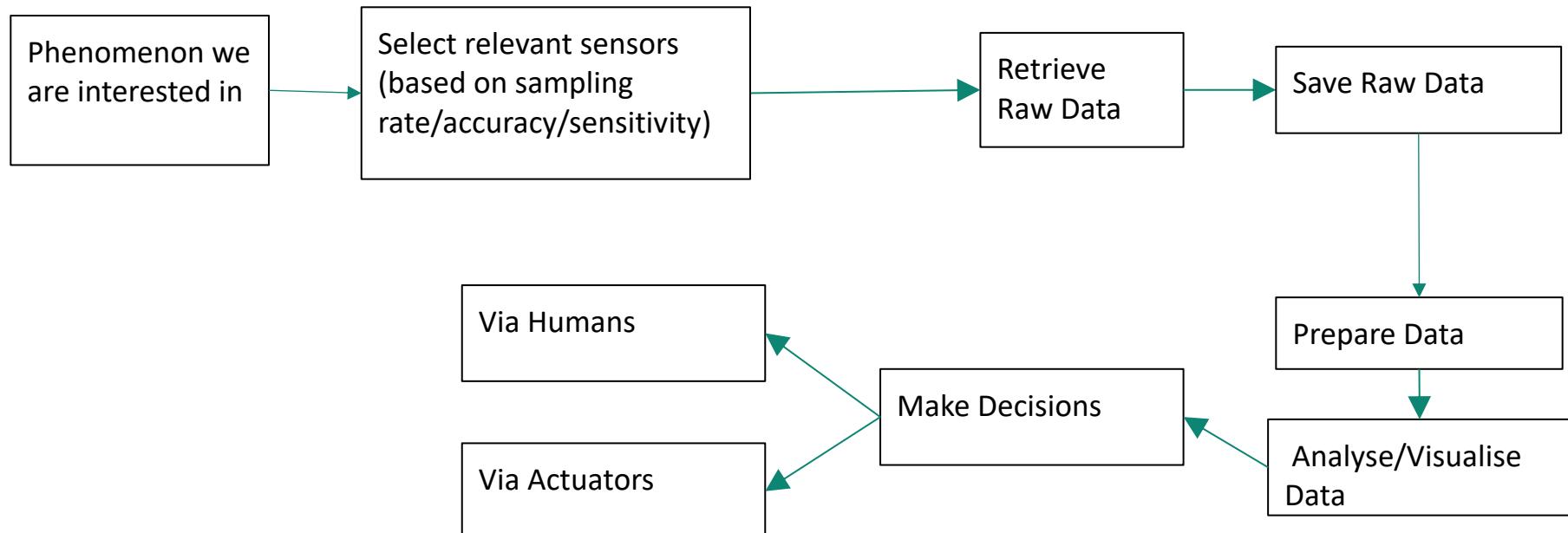
Sensor data

Formats

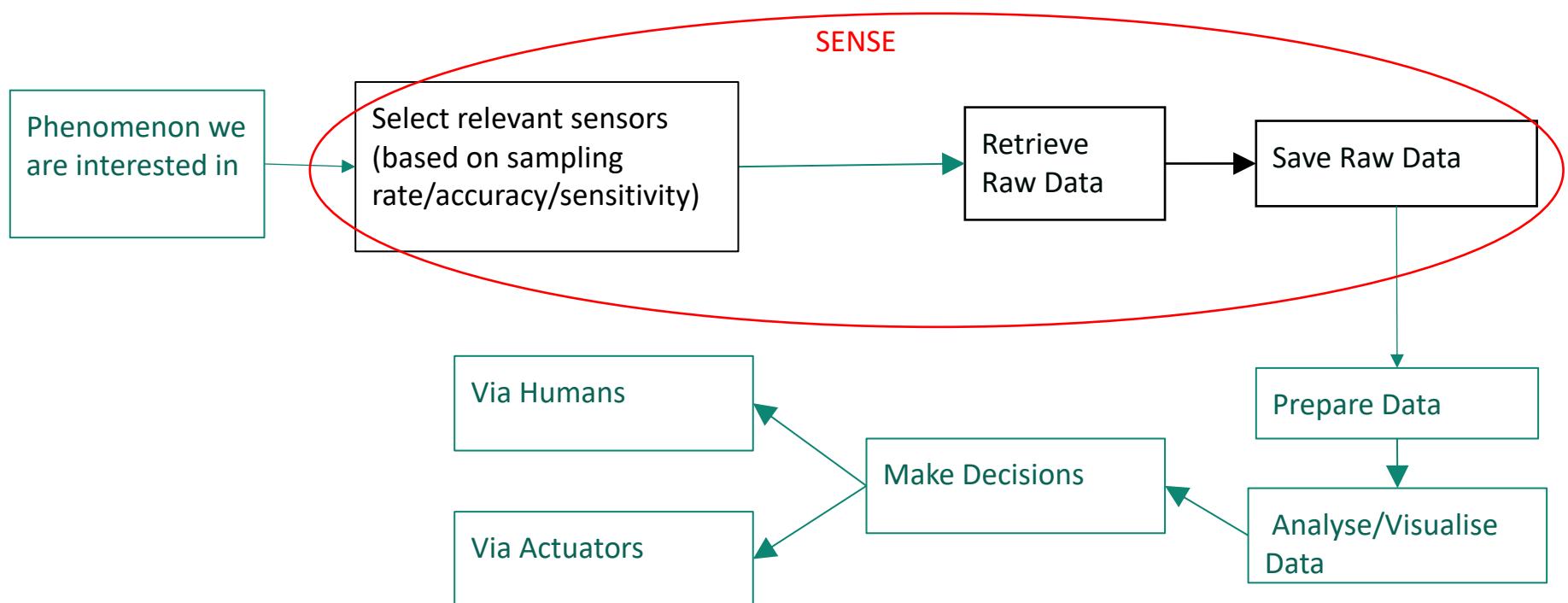
.csv files

Data preparation to clean faulty (dirty) data

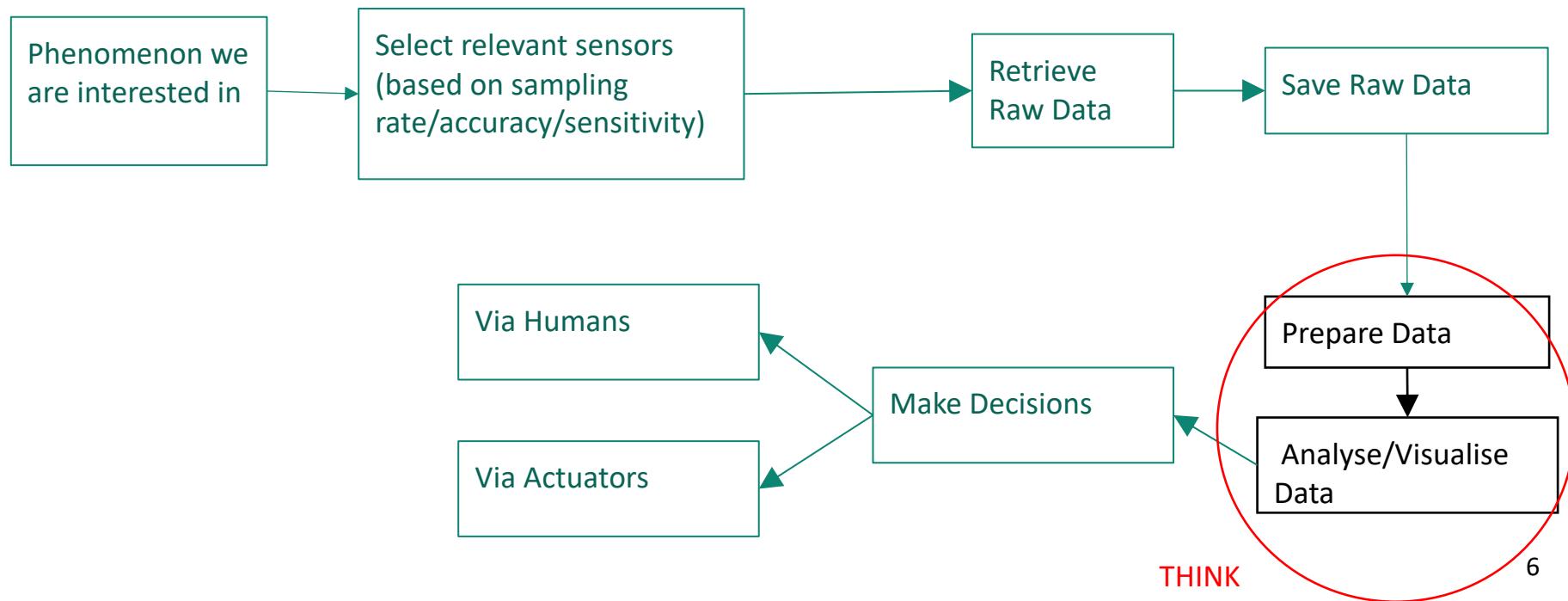
Workflow: Using captured data



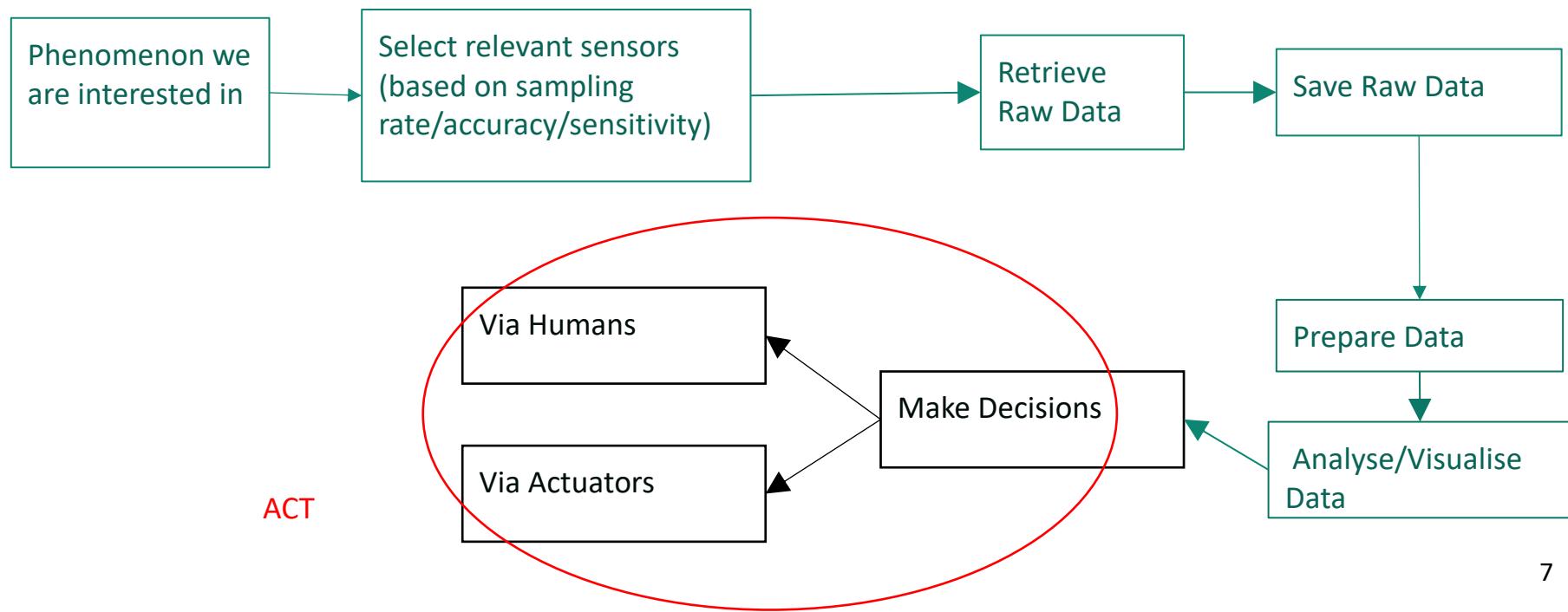
Workflow: Using captured data



Workflow: Using captured data

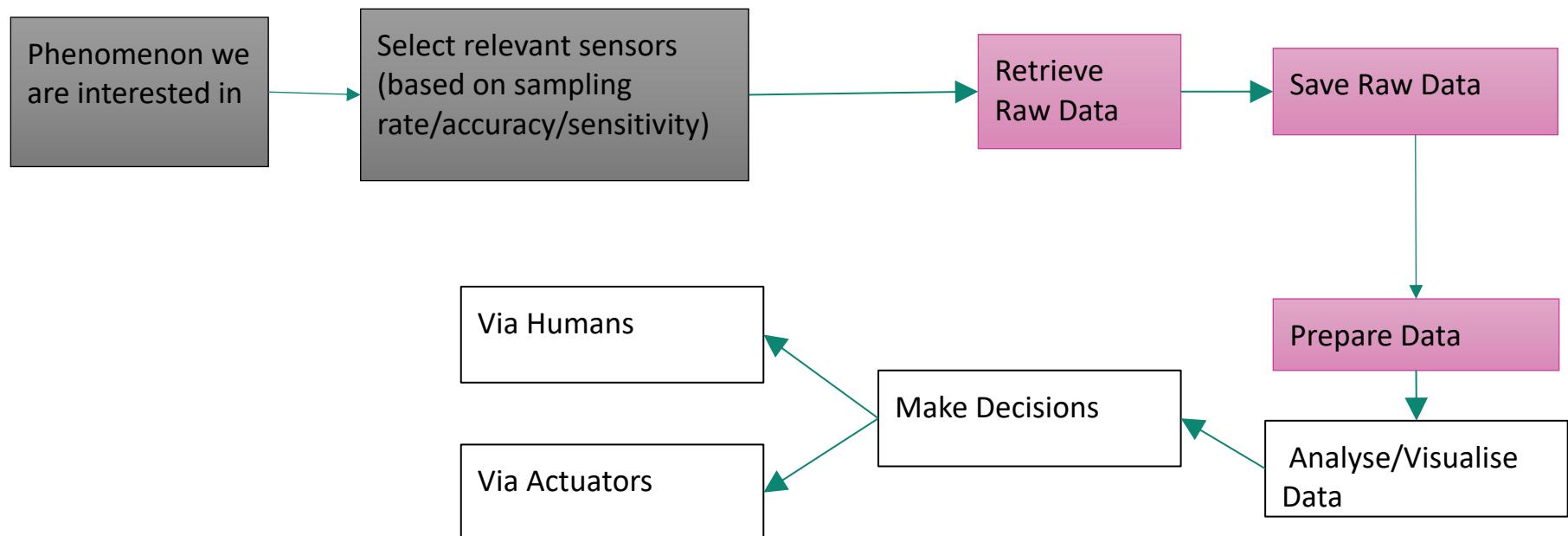


Workflow: Using captured data



ACT

Workflow: Using captured data



Saving data

During Lab Work: motion, temperature, relative humidity, soil moisture...

- Motion- Active/Inactive
 - RH &Temperature- Humidity: 35.3%, Temp: 23.50 Celsius
 - Soil moisture- range of integers from 0 to 900
- A **variety** of formats!

During Lab Work: data preparation task...

999, 1497825202, 2017/6/18 22:33:22, 61.50

- A reading every second!
 - For around 8 hours of data, the file had 27394 readings!
- High **volume** data!

Sensor data often comes in a variety of formats, and in high volume...

How did you save sensor data in your Lab work?



- How hard was that method?!

What if you wanted to collect sensor data for 24 hours??....Copying from serial monitor is not practical!

- What about date/time information for each sensor reading?

A screenshot of a Mac OS X terminal window titled '/dev/cu.usbmodem1421 (Arduino/Genuino Uno)'. The window displays a series of sensor readings: Humidity: 34.90 %, Temp: 24.30 Celsius; followed by nine identical entries: Humidity: 35.40 %, Temp: 24.30 Celsius. At the bottom of the window, there are several status indicators: a checked 'Autoscroll' checkbox, a 'No line ending' dropdown, a '9600 baud' dropdown, and a 'Clear output' button.

```
Humidity: 34.90 %, Temp: 24.30 Celsius
Humidity: 35.40 %, Temp: 24.30 Celsius
Humidity: 35.30 %, Temp: 24.30 Celsius
Humidity: 35.30 %, Temp: 24.30 Celsius
Humidity: 35.30 %, Temp: 24.30 Celsius
Humidity: 35.20 %, Temp: 24.20 Celsius
Humidity: 35.20 %, Temp: 24.20 Celsius
Humidity: 35.20 %, Temp: 24.20 Celsius
```

Data Storage, or saving data is an important step in data capture workflow

How can sensor data be saved?

- **Persistent storage**
- .csv files
- .json files
- .txt files
- Could be on the cloud, SD cards, hard drives ...



Photo by [Simson Petrol](#) on [Unsplash](#)



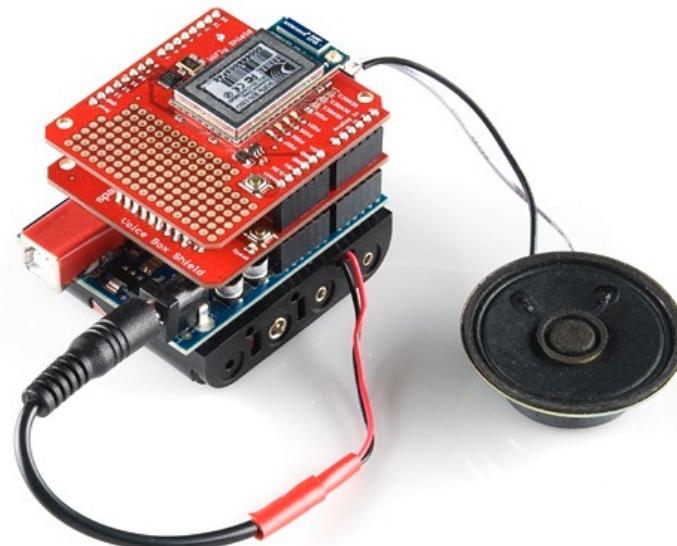
Upcoming Lab works, you will be to save sensor data to an SD card attached to the Arduino ...

We will save data to an SD card using the data logger shield in the Arduino....

What is a shield?

Arduino shields are stackable circuit boards that provide extra functionality when you stack them on top of your Arduino board.

E.g you want to connect your Arduino to internet and post on twitter? There is a shield for that!



<https://learn.sparkfun.com/tutorials/arduino-shields>

Arduino shields



Data Logger Shield (this is what we will use this week to save data)

- GPS Logger Shield
- MP3 Shield
- Bluetooth Shield
- Ethernet Shield
- WiFi Shield
- NFC/RFID Shield ...
- Many more.....



Arduino shields

Many Arduino shields are stackable. You can connect many shields together to create a "Big Mac" of Arduino modules. You could, for example, combine an Arduino Uno with a Voice Box Shield, and a WiFi Fly Shield to create a WiFi Talking Stephen Hawking(TM).

Shields are often supplied with either an example sketch, or a library. So, not only do they just simply plug into your Arduino, but all you need to do to make them work is upload up some example code to the Arduino.

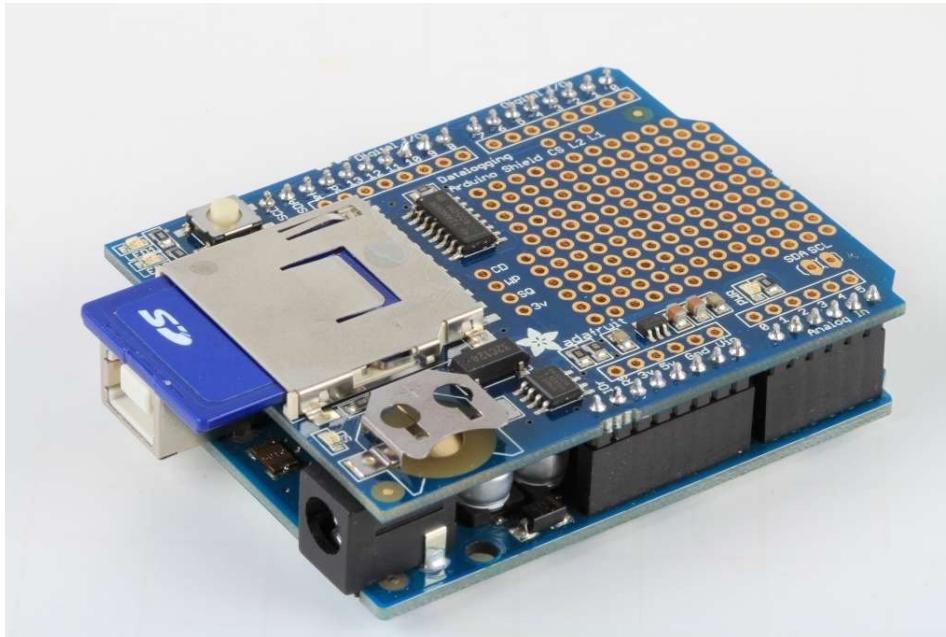


<https://learn.sparkfun.com/tutorials/arduino-shields>

Data Logger Shield



The Arduino has limited storage space, but this easy-to-use shield (along with the SD library) allows us to use an SD card for plenty of extra storage.



Data Logger Shield



Tutorial: How to Use Arduino Data Logger Shield to Save Data and Time to SD Card

<https://www.youtube.com/watch?v=wVKZpOUuCzw>

Recording date & time (timestamps)

Date/Time logs are important to provide accurate information.

Sensor data record instance observations regarding ground truth (what happened in a specific point in time in the environment)

Usually, the kind of observations we record using sensors change over time.

Eg: Temperature, Motion, Humidity, Lux

Recording date & time (timestamps)

When we do not need timestamps:

Simple If-This-Then-That scenario –

Eg: If motion is detected, switch on the lights

When we need timestamps:

If we want to find out deeply about a phenomenon

Eg:

What times of the day is the lab most active?

Should we ignore motion lasting for less than 5 seconds? Would this be more energy efficient?

When should the lights go off?



Recording date & time (timestamps) on Arduino

Use **Data logger shield with the RTC** (real time clock): You will do this in Lab work!

- Needs a coin cell battery
- The RTC chip is a specialized chip that just keeps track of time. It can count leap-years and knows how many days are in a month

Why?

- Without the Datalogger shield with the RTC, the Arduino board doesn't know its 'Tuesday' or 'March 8th'. It only has a built-in timekeeper called millis() that can only keep track of time since the Arduino was last powered.

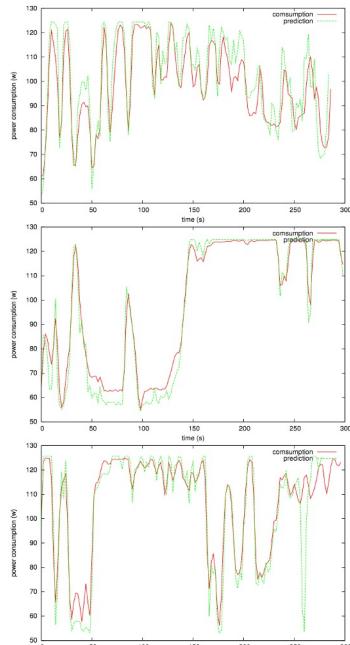
Analysing Sensor Data



Let's Look at Some Examples...

Smart Meter Data Can Identify What You Were Watching on TV...

U. Greveler, B. Justus and
 D. Loehr. "Multimedia
 content identification
 through smart meter power
 usage profiles"
*Computers, Privacy and
 Data Protection, 2012.*



Power consumption
 patterns for :

Star Trek 2011 (movie)

Star Trek TNG (S1,
 episode1)

Body of Lies (movie)

Fig. 5. power prediction vs. consumption: first 5 minutes of the movie Star Trek 11 (top), of episode 1, Star Trek TNG season 1, of the movie Body of Lies (bottom)

Sensor Data from a House ...



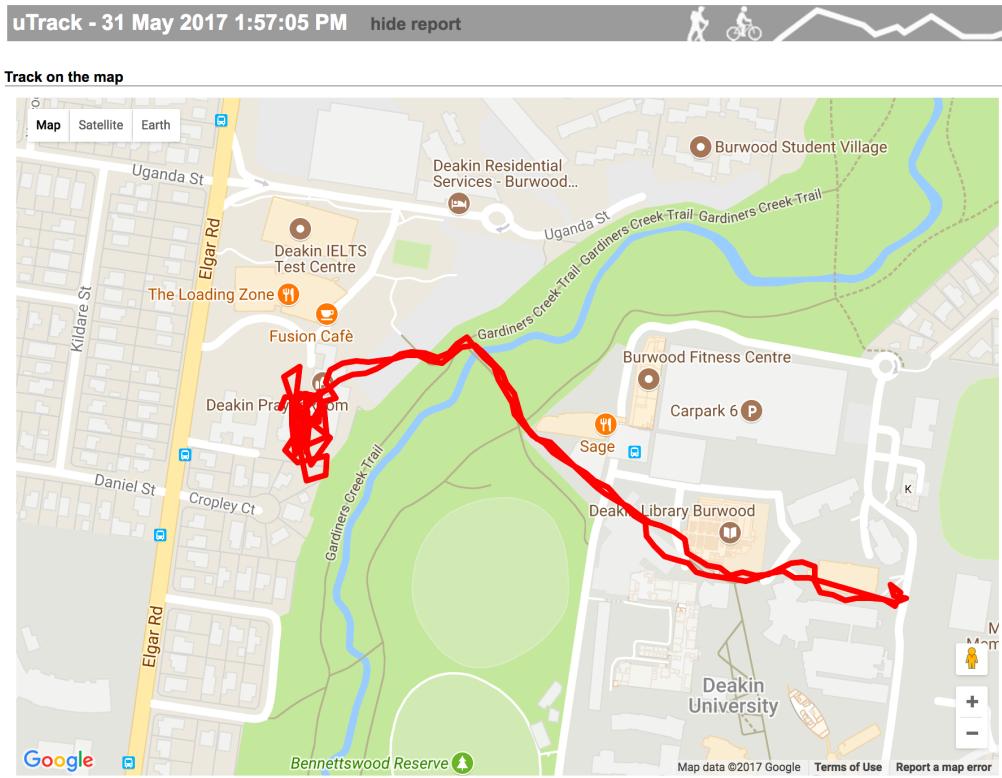
```
6024, "component_name": "motion", "value": "inactive", "occurred_at": "2016-08-01T11:11:15.357Z", "created_at": "2016-08-01T11:11:17.000Z", "updated_at": "2016-08-01T11:11:17.666Z", "device_id": 13}, {"id": 6025, "component_name": "motion", "value": "inactive", "occurred_at": "2016-08-01T11:11:16.330Z", "created_at": "2016-08-01T11:11:18.555Z", "updated_at": "2016-08-01T11:11:18.555Z", "device_id": 14}, {"id": 6026, "component_name": "motion", "value": "active", "occurred_at": "2016-08-01T11:11:19.569Z", "created_at": "2016-08-01T11:11:21.538Z", "updated_at": "2016-08-01T11:11:21.538Z", "device_id": 14}, {"id": 6027, "component_name": "motion", "value": "inactive", "occurred_at": "2016-08-01T11:11:31.285Z", "created_at": "2016-08-01T11:11:33.277Z", "updated_at": "2016-08-01T11:11:33.277Z", "device_id": 16}, {"id": 6028, "component_name": "motion", "value": "active", "occurred_at": "2016-08-01T11:11:46.188Z", "created_at": "2016-08-01T11:11:48.449Z", "updated_at": "2016-08-01T11:11:48.449Z", "device_id": 14}, {"id": 6029, "component_name": "motion", "value": "active", "occurred_at": "2016-08-01T11:11:46.422Z", "created_at": "2016-08-01T11:11:48.640Z", "updated_at": "2016-08-01T11:11:48.640Z", "device_id": 6}, {"id": 6030, "component_name": "motion", "value": "active", "occurred_at": "2016-08-01T11:11:48.121Z", "created_at": "2016-08-01T11:11:50.238Z", "updated_at": "2016-08-01T11:11:50.238Z", "device_id": 4}, {"id": 6031, "component_name": "motion", "value": "active", "occurred_at": "2016-08-01T11:11:48.186Z", "created_at": "2016-08-01T11:11:50.638Z", "updated_at": "2016-08-01T11:11:50.638Z", "device_id": 13}, {"id": 6032, "component_name": "motion", "value": "inactive", "occurred_at": "2016-08-01T11:11:56.366Z", "created_at": "2016-08-01T11:11:58.321Z", "updated_at": "2016-08-01T11:11:58.321Z", "device_id": 14}, {"id": 6033, "component_name": "motion", "value": "inactive", "occurred_at": "2016-08-01T11:11:56.829Z", "created_at": "2016-08-01T11:11:59.196Z", "updated_at": "2016-08-01T11:11:59.196Z", "device_id": 6}, {"id": 6034, "component_name": "motion", "value": "inactive", "occurred_at": "2016-08-01T11:11:57.209Z", "created_at": "2016-08-01T11:11:59.219Z", "updated_at": "2016-08-01T11:11:59.219Z", "device_id": 9}, {"id": 6035, "component_name": "motion", "value": "active", "occurred_at": "2016-08-01T11:12:25.352Z", "created_at": "2016-08-01T11:12:28.007Z", "updated_at": "2016-08-01T11:12:28.007Z", "device_id": 6}, {"id": 6036, "component_name": "motion", "value": "active", "occurred_at": "2016-08-01T11:12:33.546Z", "created_at": "2016-08-01T11:12:35.998Z", "updated_at": "2016-08-01T11:12:35.998Z", "device_id": 9}, {"id": 6037, "component_name": "contact", "value": "open", "occurred_at": "2016-08-01T11:12:36.573Z", "created_at": "2016-08-01T11:12:38.635Z", "updated_at": "2016-08-01T11:12:38.635Z", "device_id": 3}, {"id": 6038, "component_name": "acceleration", "value": "active", "occurred_at": "2016-08-01T11:12:36.695Z", "created_at": "2016-08-01T11:12:38.778Z", "updated_at": "2016-08-01T11:12:38.778Z", "device_id": 3}, {"id": 6039, "component_name": "acceleration", "value": "inactive", "occurred_at": "2016-08-01T11:12:40.299Z", "created_at": "2016-08-01T11:12:42.422Z", "updated_at": "2016-08-01T11:12:42.402Z", "device_id": 3}, {"id": 6040, "component_name": "motion", "value": "inactive", "occurred_at": "2016-08-01T11:12:42.735Z", "created_at": "2016-08-01T11:12:44.739Z", "updated_at": "2016-08-01T11:12:44.739Z", "device_id": 1}
```

-> house was unoccupied from 7 am - 6 pm

-> occupants went to sleep by 11 pm

-> first activity in the morning recorded at 6.15 am

Location data from my phone



31_May_20171_57_05_PM.gpx

```

<trkpt lat="-37.845943" lon="145.110718">
  <ele>80</ele>
  <time>2017-05-31T03:57:28.112Z</time>
</extensions>
</trkpt>
<trkpt lat="-37.846081" lon="145.110626">
  <ele>103</ele>
  <time>2017-05-31T03:57:27.112Z</time>
</extensions>
<geotracker:meta c="341.5" s="0.52" />
</trkpt>
<trkpt lat="-37.846104" lon="145.110779">
  <ele>46</ele>
  <time>2017-05-31T03:57:25.106Z</time>
</extensions>
<geotracker:meta c="341.5" s="0.33" />
</trkpt>
<trkpt lat="-37.846138" lon="145.110672">
  <ele>140</ele>
  <time>2017-05-31T03:57:24.100Z</time>
</extensions>
<geotracker:meta c="341.5" s="0.68" />
</trkpt>
<trkpt lat="-37.846138" lon="145.110672">
  <ele>140</ele>
  <time>2017-05-31T03:57:22.090Z</time>
</extensions>
<geotracker:meta c="341.5" s="0.7" />
</trkpt>

```

A document describing the data

Mention if there are missing values

Clarify the format

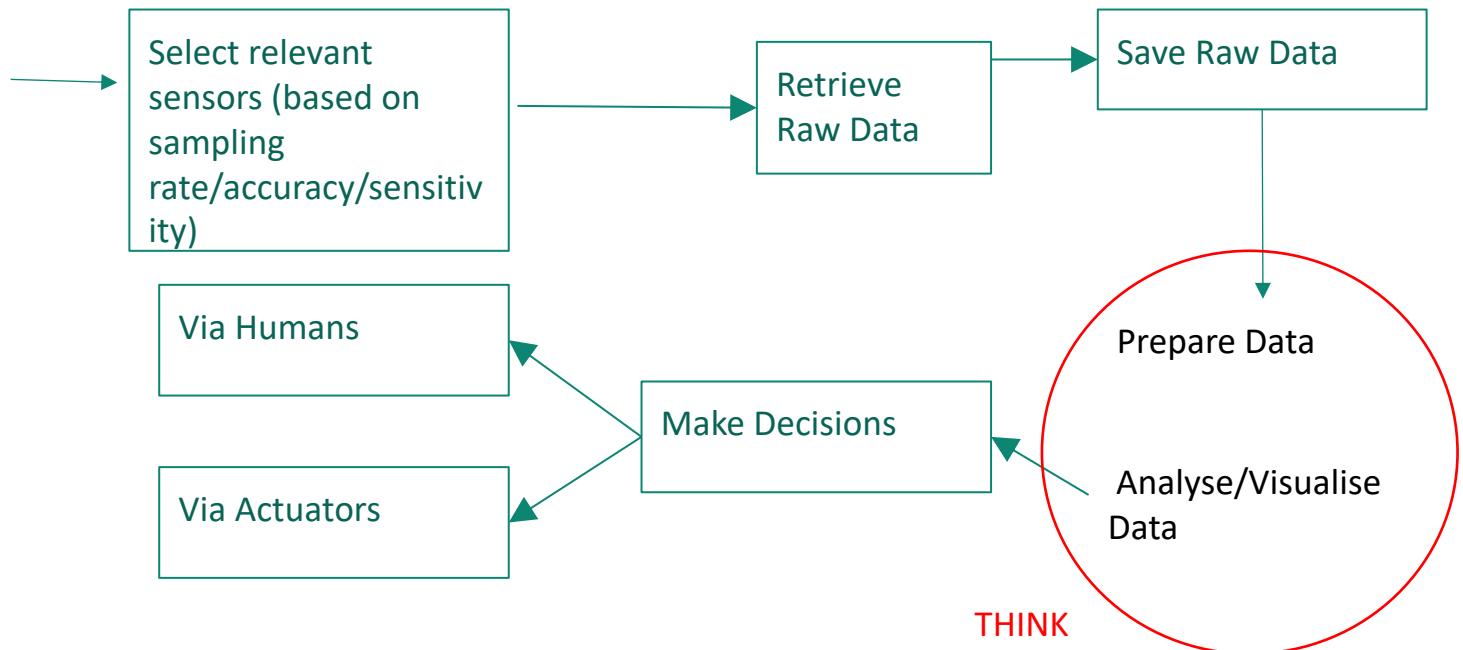
What values are recorded

How to analyse sensor data



Step1: Prepare (clean) the data

Step 2: Run functions/algorithms (Excel, Python, Matlab, R, Java, JavaScript, C, C++ ,)



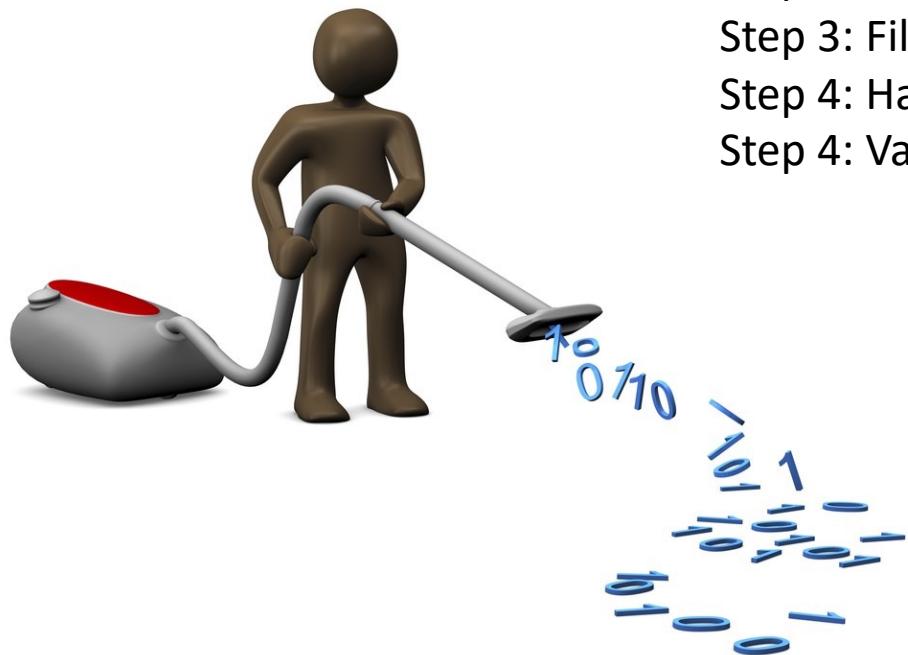
Data Cleaning



Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

<https://www.tableau.com/learn/articles/what-is-data-cleaning>

- Step 1: Remove duplicate or irrelevant observations
- Step 2: Fix structural errors
- Step 3: Filter unwanted outliers
- Step 4: Handle missing data
- Step 4: Validate and QA

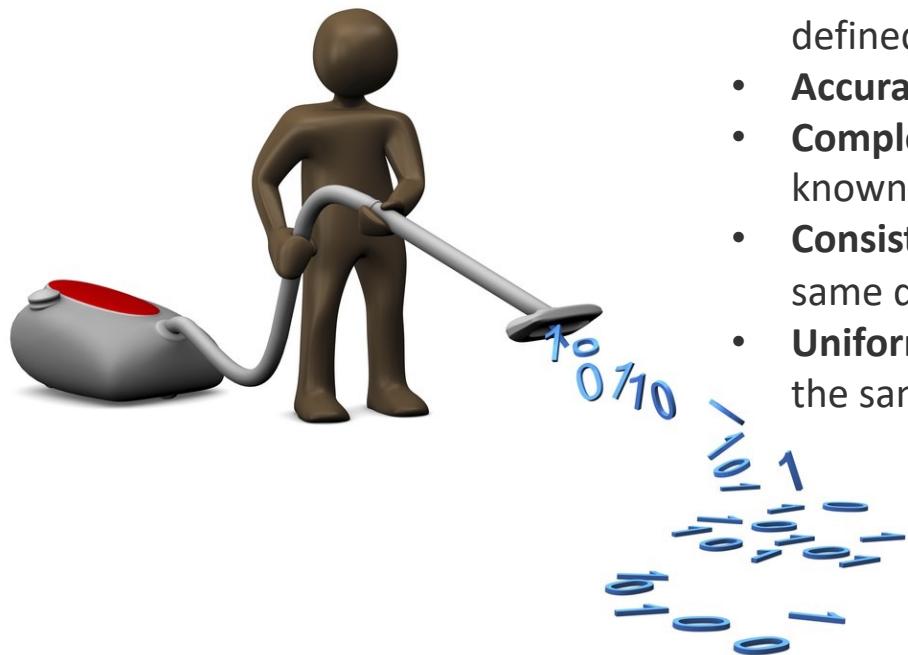


Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

<https://www.tableau.com/learn/articles/what-is-data-cleaning>

5 characteristics of quality data

- **Validity.** The degree to which your data conforms to defined problem scenario or constraints.
- **Accuracy.** Ensure your data is close to the true values.
- **Completeness.** The degree to which all required data is known.
- **Consistency.** Ensure your data is consistent within the same dataset and/or across multiple data sets.
- **Uniformity.** The degree to which the data is specified using the same unit of measure.



Simple ways to analyse sensor data

- **Descriptive statistics**

Mean, median, mode, standard deviation...

(We will do this in Lab works)

Eg: The mean temperature of the lab is 23.5 C, over a period of 24 hours

The standard deviation of the lab temperature is 2 C

-> ***What does this tell us?***

Simple ways to analyse sensor data

- **Descriptive statistics**

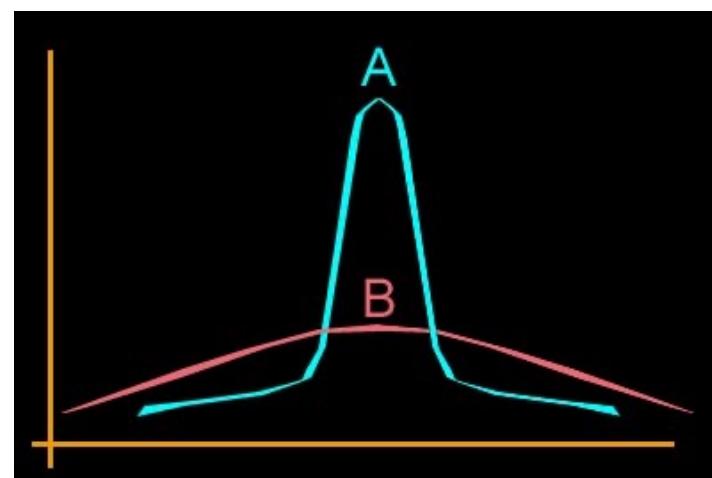
Mean, median, mode, standard deviation...

The average (mean) daily temperature in Hawaii is 78 degrees Fahrenheit. The average daily temperature in Oklahoma is 77 degrees Fahrenheit.

But there is a difference in standard deviation. Hawaii is a mere 2.52 while Oklahoma came in at 10.57.

A=?

B=?

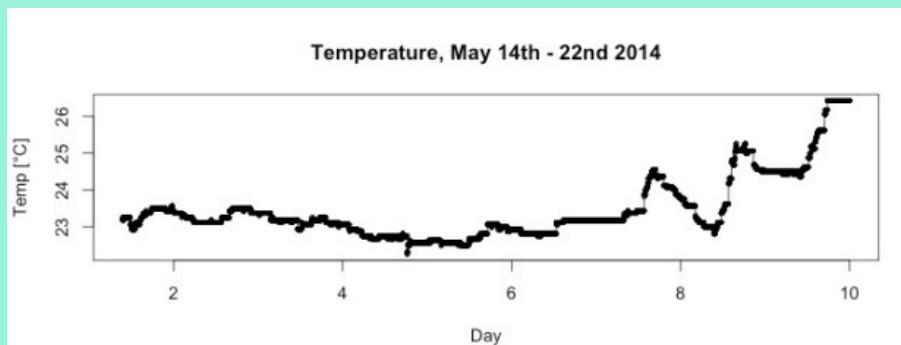


Hawaii	Oklahoma
72	56
75	61
76	66
77	69
77	74
78	76
78	76
78	77
78	77
78	77
78	78
78	78
79	80
79	85
80	88
81	93
84	98

Simple ways to analyse sensor data

- Plot the data and look at the distribution

Eg: The temperature data on the minute level from May 14th until May 22nd in southern Germany is plotted below-



→ **What does this tell us?**

(Note: We will use visualization methods like this in the coming weeks)

<https://dzone.com/articles/processing-and-analyzing>

Today we talked about...



Saving data
Analyzing data



Final Project

Time to start thinking about your project ideas

Project Report is due in Week 10.

Project Poster and Presentation video are due in Week 11.



Final Project

Home Content Discussions Assessment ▾ Tools ▾ Setup 2022 T2

Overview

Bookmarks

Site Schedule 1

Table of Contents 117

Unit Information 16

Assessment Resources 10

Lab Reports 4

Project 6

Learning Resources 72

Online Classrooms 6

Class Recordings 1

Project ▾

Print

Add dates and restrictions...

This is an individual assessment.

The project in SIT123 is worth 50% of the total grading. This will include Project report on the idea, research and coding [40%], the project poster [5%] and 1-2 minutes video presentation [5%].

The project Report is due on Friday 8pm in Week 10, whereas the poster and video presentations are due on Friday 8pm in Week 11.

Project Description File is available. You can start thinking on the idea for your final project as early as you can.

STUDENT DECLARATION is MANDATORY for all students to sign and attach to their project reproto.

Upload / Create ▾

Existing Activities ▾

Bulk Edit

SIT123 Project [T2 2022]

Word Document

Student Declaration

Word Document

Rubric SIT123 Final Project-Poster-Presentation

PDF document



Final Project

SIT123 Data Capture Technologies, T2 2022

Student Declaration – Final Project

Plagiarism is a serious academic misconduct. It means when you use someone else's intellectual output (text, idea, images etc.) and present them (without appropriate and standard way of acknowledging the source) as your own.

PLEASE READ THE FOLLOWING STATEMENTS AND SIGN BELOW TO INDICATE THAT YOU HAVE UNDERSTOOD THE STATEMENTS:

- I declare that this assessment is my own work and demonstrates my own abilities and knowledge and does not involve plagiarism.
- I have taken serious precautions to make sure other students do not copy my work.
- I declare that I have properly provided all the citations and references to the original sources when I had to present ideas that are not my own.
- I understand very clearly that plagiarism can result in penalty of zero marks for the assessment.
- I understand that serious offences of plagiarism (e.g. submitting work of students from previous semester, or hired services work) will be forwarded to the faculty for further actions.

Student name: <<Student name>>

Signature:

Date:

<<After signing attach this to your 'Final Project Report' before submission. This declaration is mandatory to be signed by all students>>

Remember....

During the practical sessions, you can do the practicals as a group, **but you must answer the questions individually!**

You can help others, but do not just give your answer/code. This does not help your friend to learn, and **may result in plagiarism issues for yourself and the other student(s).**

Note: Sharing collected sensor data & screen shots is fine.

References

<https://learn.adafruit.com/adafruit-data-logger-shield/using-the-sd-card>

<https://learn.sparkfun.com/tutorials/arduino-shields>

<https://learn.adafruit.com/adafruit-data-logger-shield/using-the-real-time-clock>

<http://www.dummies.com/education/math/statistics/how-to-interpret-standard-deviation-in-a-statistical-data-set/>

<https://dzone.com/articles/processing-and-analyzing>

Any Questions!

