



**M. DIAMANT, A. VUOSKU, M.
UUSITALO, M. HUSSAIN**

Energy consumption in Oulu

Seminar 2

**How do the buildings' characteristics
impact their yearly energy consumption
and can we cluster some different
consumption profiles?**

A large, abstract graphic on the left side of the slide features several overlapping geometric shapes. It includes a dark grey triangle at the top left, a white triangle below it, and a large black triangle with a diagonal hatching pattern. A thin white line runs diagonally across the black triangle. At the bottom right, there is a solid orange-red rectangle.

The Data set

Two documents

Metadata:

- Building's specific information, e.g.: floorcount, gross area, year of construction

Records:

- Buildings' hourly energy consumption records
- Heat and Electricity

The Metadata

	property_id	property_name	district_key	district_name	property_address	postal_code	postal_area	grossarea	totalgrossarea	volume	totalfloorare
0	675201	Ylikiimingin koulun yläkoulu	16	Ylikiiminki	Harjutie 17 / Opinkuja 4	91300	YLIKIIMINKI	2079.0	2345.0	8540.0	2439.
1	675701	Ylikiimingin terveysasema	16	Ylikiiminki	Vesaisenlinnantie 2	91300	YLIKIIMINKI	844.0	975.0	3340.0	886.
2	656401	Martinniemen koulu ja kirjasto	20	Haukipudas	Jokisuuntie 1	90850	MARTINNIEMI	3018.0	3415.0	14773.0	3298.

The Consumption records

	property_id	property_internal_id	property_name	consumption_measure	year	month	day	starting_hour	consumption	keyfield
0	656601	{267DD242-39FA-48B1-B4F6-8F88F77D9081}	Pihlajakoti	Lämpö	2018	12	20	0	17	146671951
1	656601	{267DD242-39FA-48B1-B4F6-8F88F77D9081}	Pihlajakoti	Lämpö	2018	12	16	0	14	146671952
2	656601	{267DD242-39FA-48B1-B4F6-8F88F77D9081}	Pihlajakoti	Lämpö	2018	12	9	0	15	146672451

Data Cleaning

A lot of missing, redundant,
untranslated data
or non-optimal formats

Missing Data

**Out of the 536 buildings, only
286 have actual records**



Data Transformation

Metadata

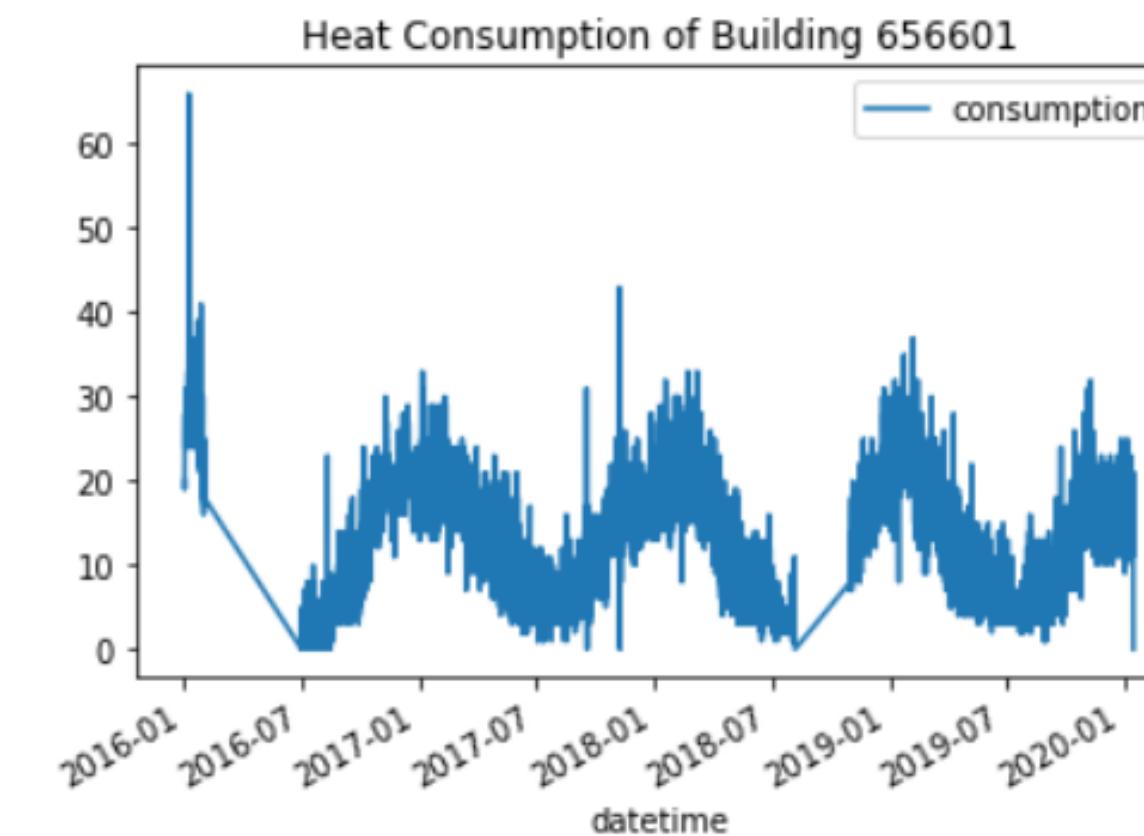
Translation (intended uses)
Other issues are only manageable downstream

Records

Translation (Energy type)
Reformat date time
Time window narrowing

Updated Consumption records

	property_id	property_internal_id	property_name	consumption_measure	consumption	keyfield	datetime
0	656601	{267DD242-39FA-48B1-B4F6-8F88F77D9081}	Pihlajakoti	Heat	8.0	7187764	2016-09-28 14:00:00
1	656601	{267DD242-39FA-48B1-B4F6-8F88F77D9081}	Pihlajakoti	Heat	9.0	7188240	2016-09-24 14:00:00
2	656601	{267DD242-39FA-48B1-B4F6-8F88F77D9081}	Pihlajakoti	Heat	5.0	7188241	2016-09-25 14:00:00
3	656601	{267DD242-39FA-48B1-B4F6-8F88F77D9081}	Pihlajakoti	Heat	5.0	7188242	2016-09-06 14:00:00





Feature extraction

Different feature types

- Categorical features: Intended use, Postal code
District name
- Semi-continuous features: Gross area, Volume
- Discrete features (in-between): Floor count, Year
of construction

Categorical features

- List of characteristic categories

```
['511 Educational building buildings', '231 Kindergartens', '214 Health Centers']
```

- Average energy consumption by building

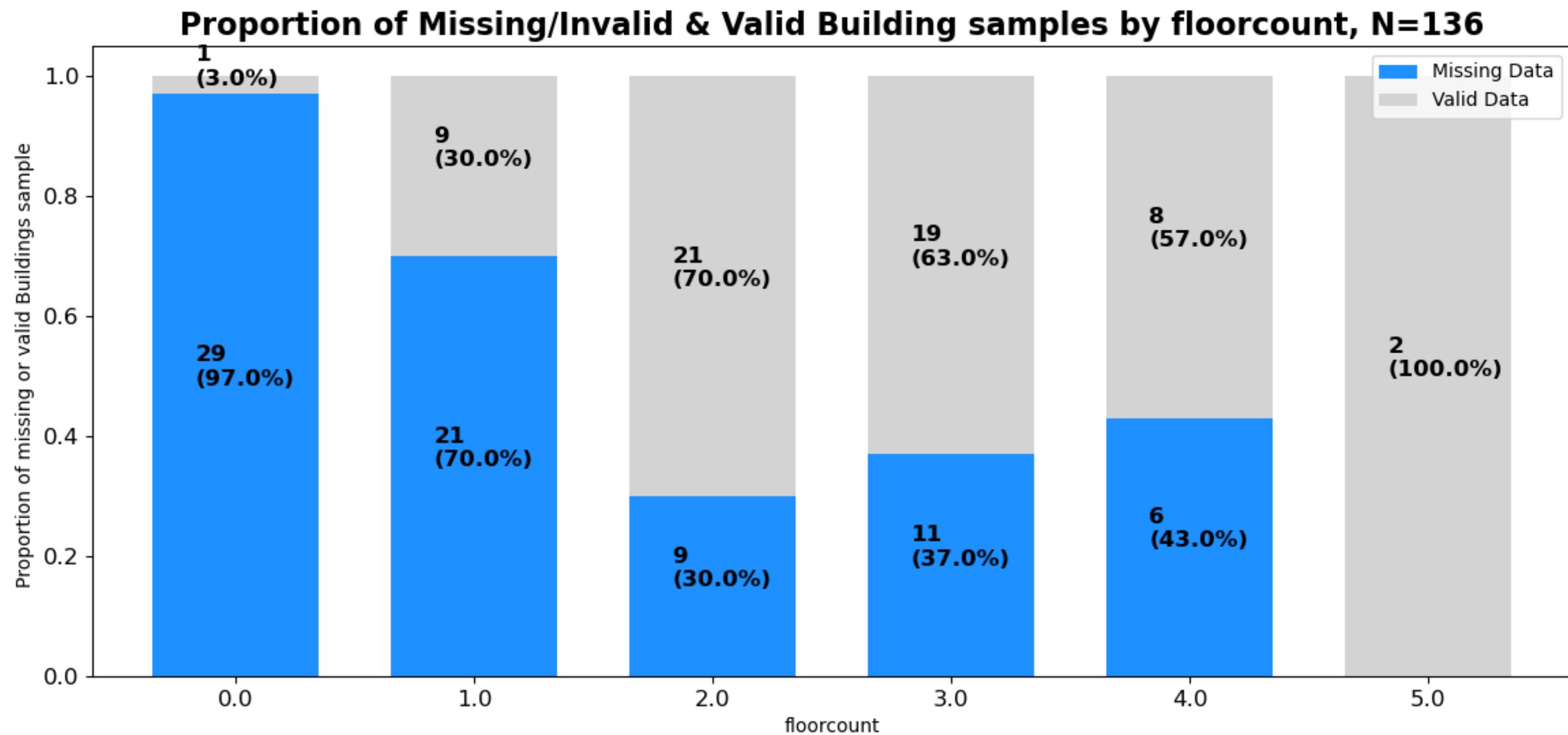
2	511 Educational building buildings	656301	NaN	26.44
3	511 Educational building buildings	677302	NaN	NaN
4	511 Educational building buildings	657302	NaN	NaN
5	231 Kindergartens	688202	NaN	NaN
6	231 Kindergartens	646101	32.77	13.63

- Average consumption group by category

	intended_use	average_heat	average_electricity
0	231 Kindergartens	32.77	12.35
1	511 Educational building buildings	NaN	31.29
2	999 Buildings nec	30.52	12.41

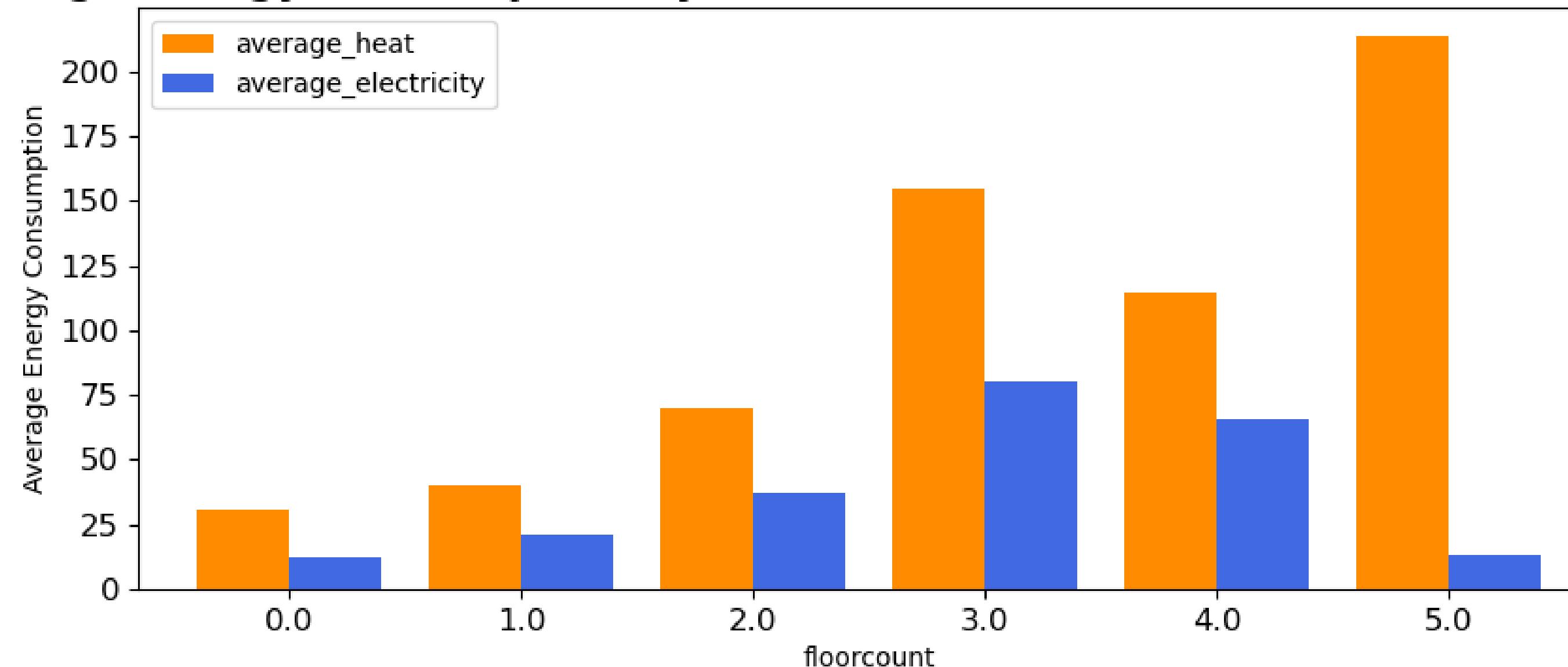


Missing and invalid data



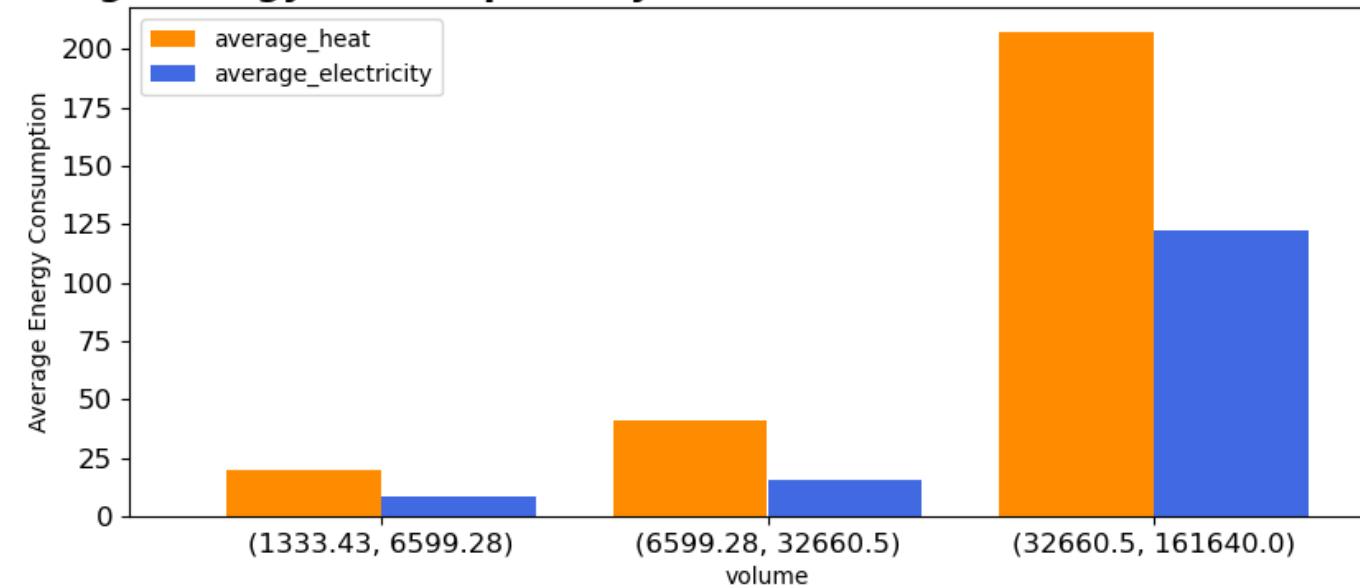
Average Consumption visualisation

Average Energy Consumption by floorcount, N=60(N: number of valid samples)

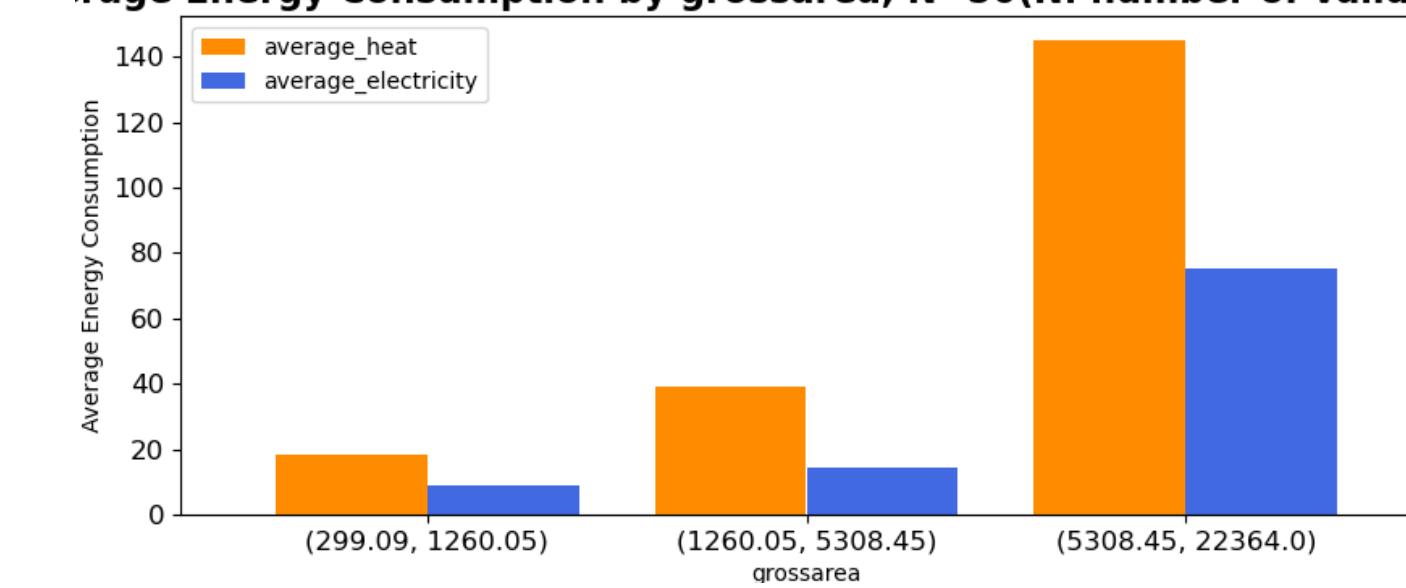


Some interesting graphs

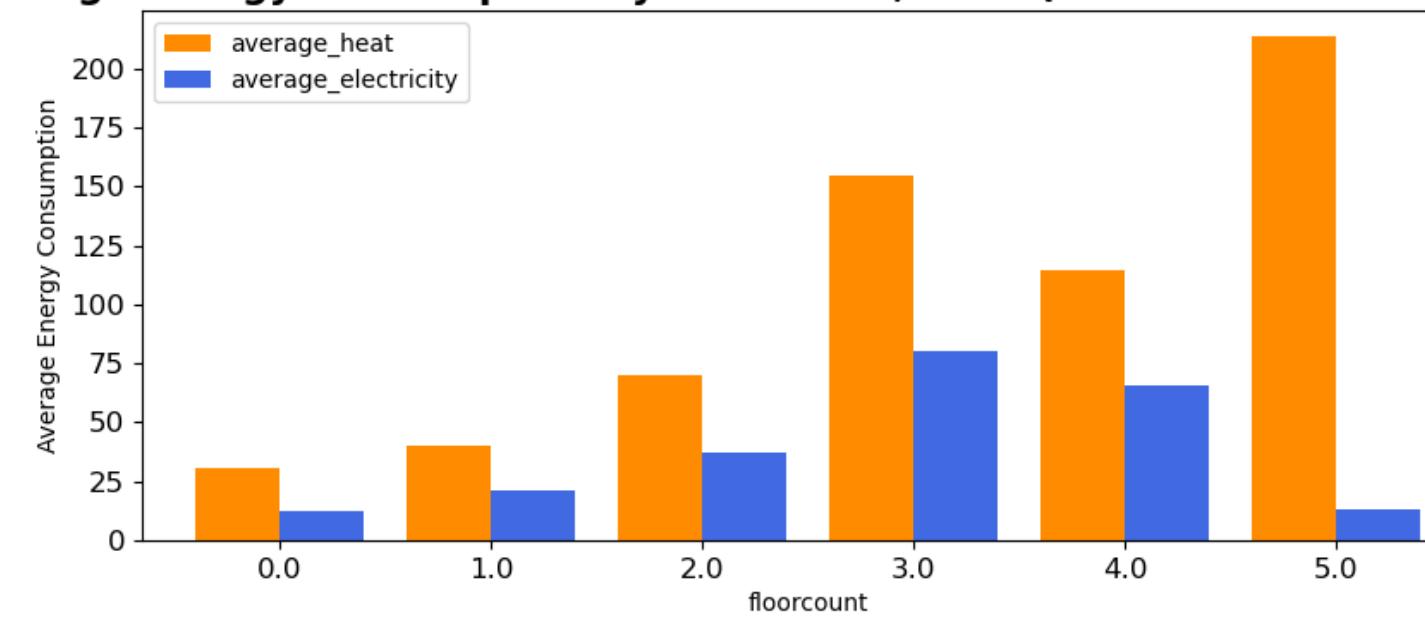
Average Energy Consumption by volume, N=44(N: number of valid samples)



Average Energy Consumption by grossarea, N=50(N: number of valid samples)

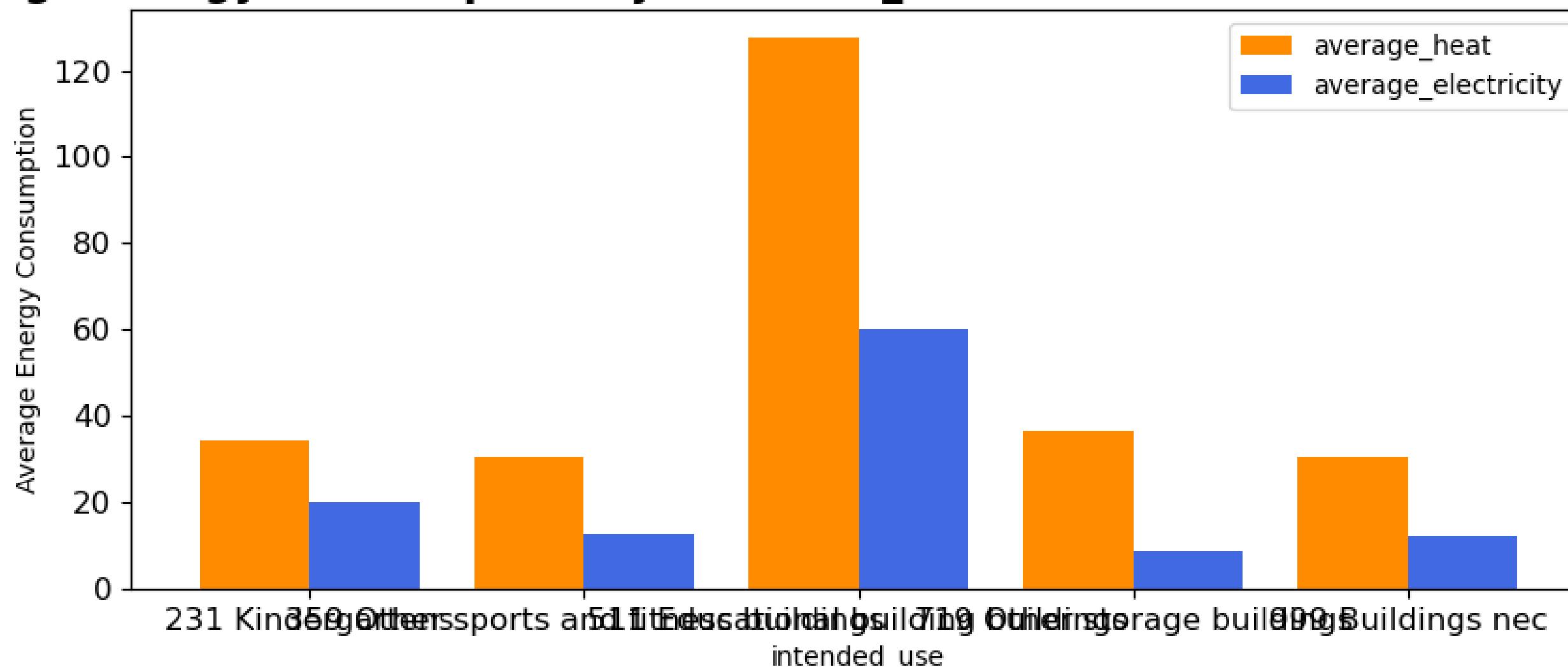


Average Energy Consumption by floorcount, N=60(N: number of valid samples)



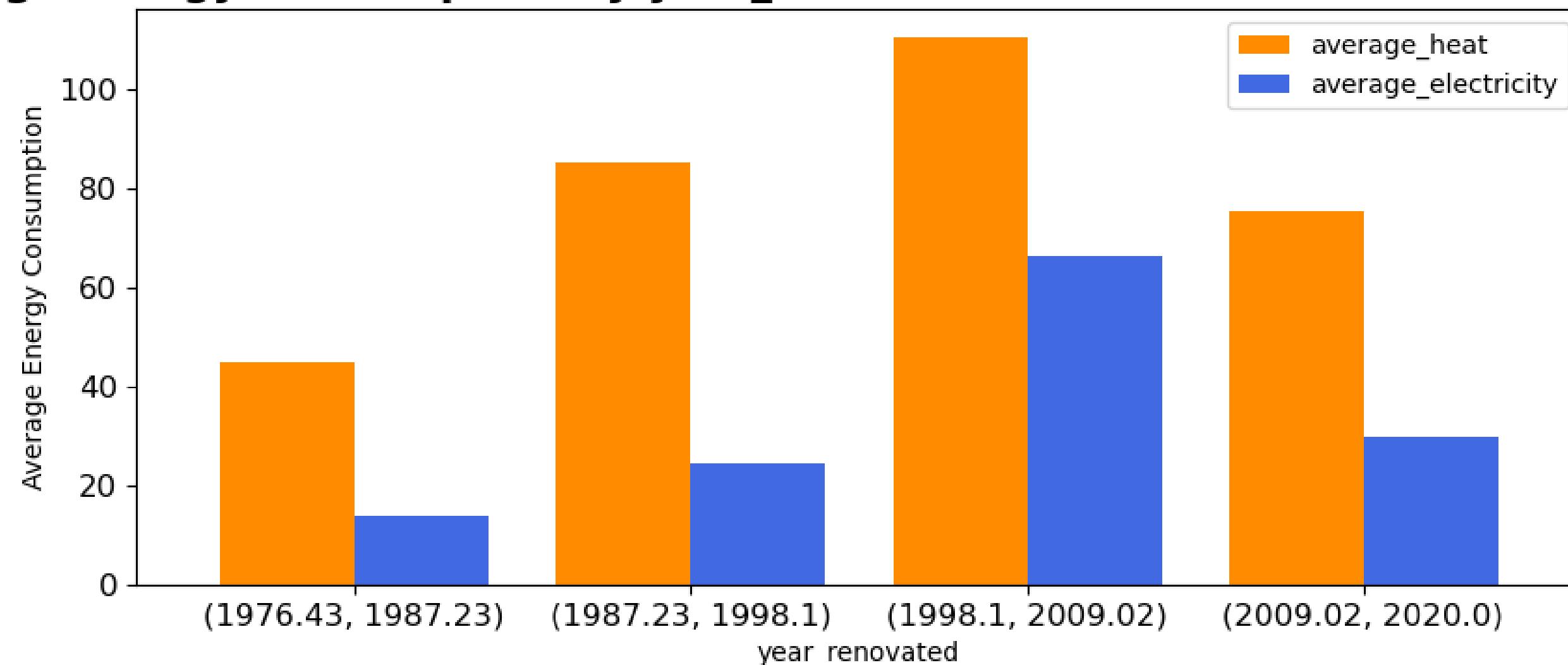
Some interesting graphs

Average Energy Consumption by intended_use, N=22(N: number of valid samples)



Some interesting graphs

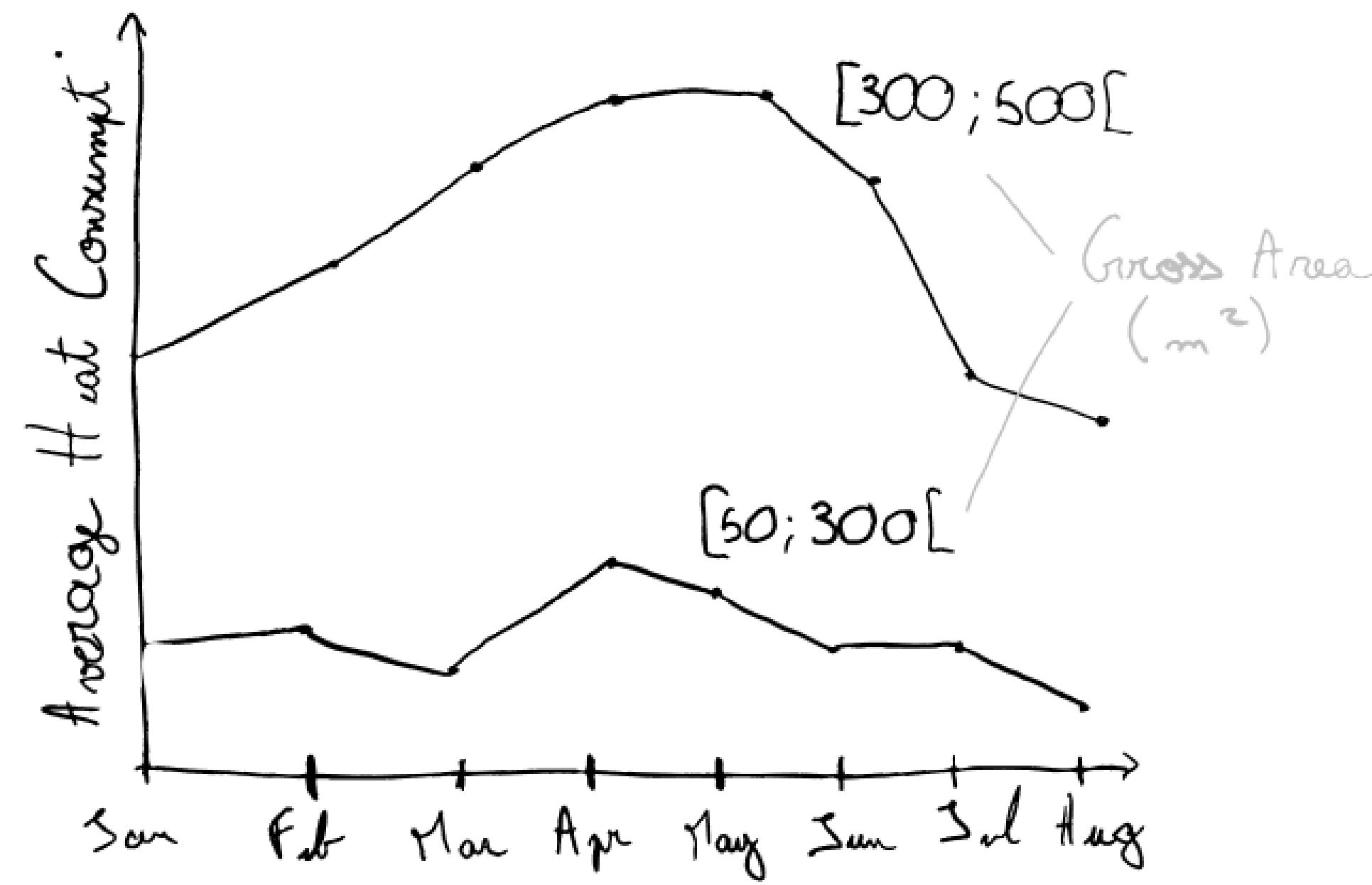
Average Energy Consumption by year_renovated, N=56(N: number of valid samples)



Next steps

- Study semi-continuous and discrete data
- Narrow down consumption average to a monthly or yearly average
- Compare characteristics correlated impact

Example: Monthly Average Heat consumption by gross area



About the group work

Distributed work in sub-questions

Marius:

- Data collection, cleaning and transformation
- Feature extraction (categorical and semi-continuous)
- Communication and presentation creation

Arttu:

- Analysis of building's consumption through the city
- Creation of an interactive map

Mussadaq:

- Clusterization of consumption profiles
- Starting by clustering energy consumption daily profiles

Mari:

- Writing the report
- Related work researches

Conclusion

01

A complex and promising dataset

02

A lot of data cleaning and transformation

03

Some characteristics have more impact than others

04

We can already see correlation patterns between characteristics

Any questions?

