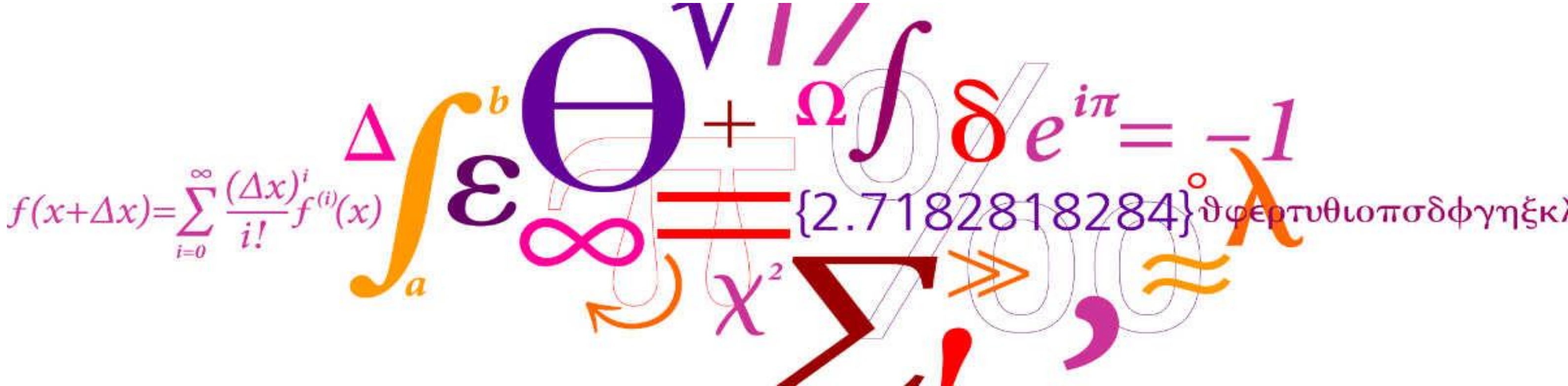
UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

Shape-based clustering of ECG time series with Dynamic Time Warping

KU105 Advanced Physiological Modeling

Final presentation

May 29th, 2018



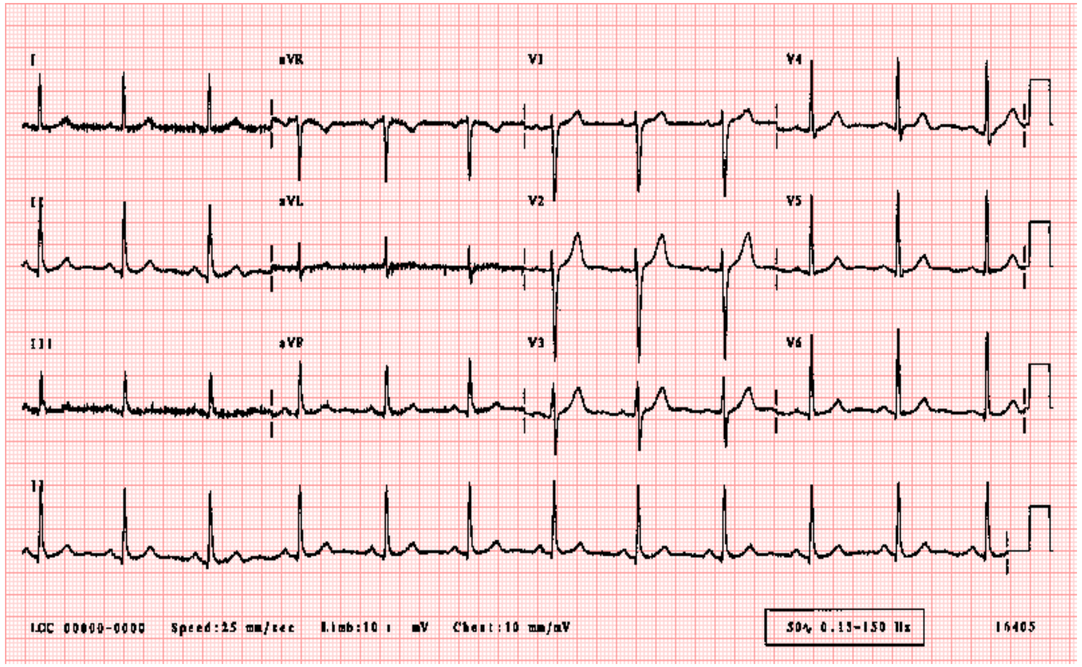
Alessandro Montemurro – DTU Compute

Today’s agenda

- Electrocardiography
- Data
- Hierarchical Clusering
- Dynamic Time Warping
- Results

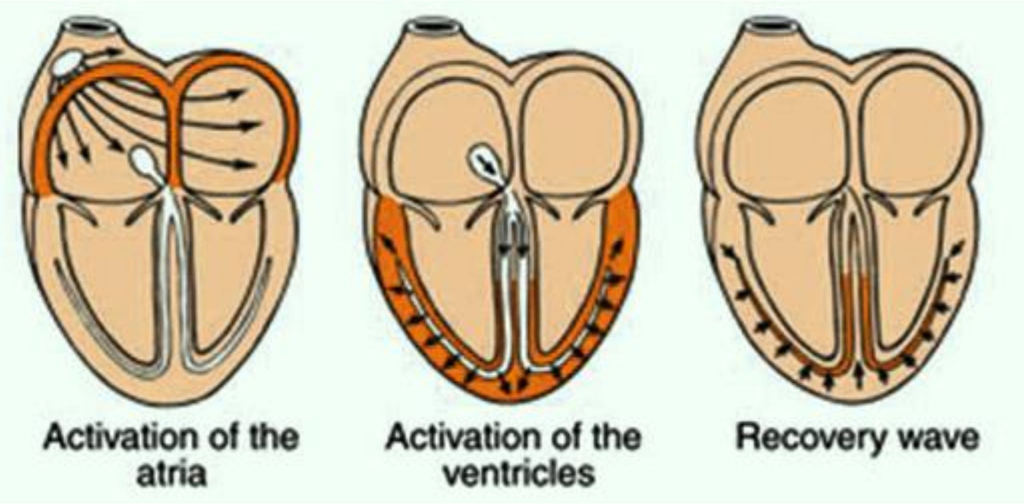
ELECTROCARDIOGRAM

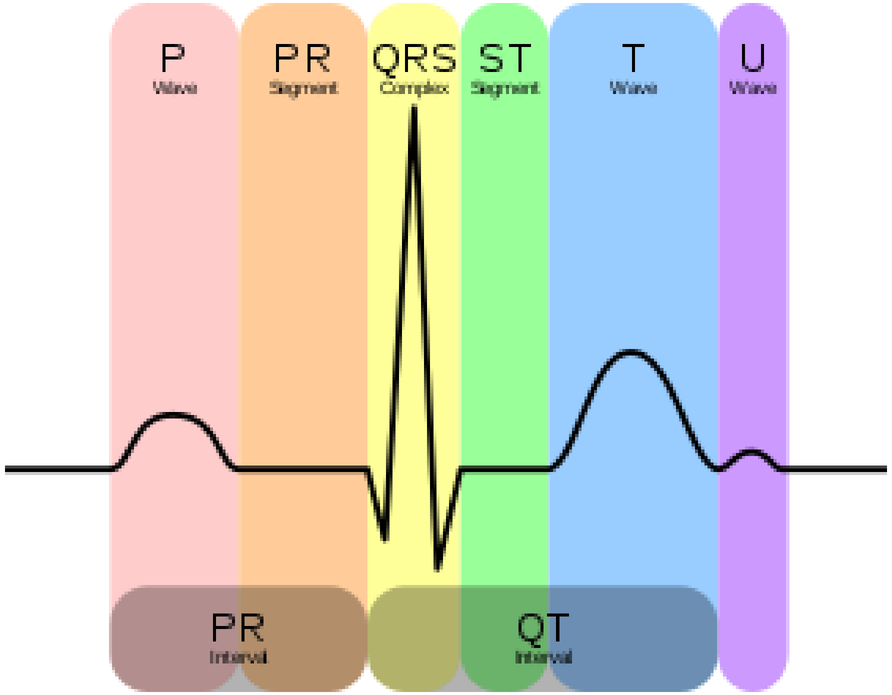
- Recording of the electrical activity of the heart
- 10 electrodes on the chest and limbs
- Measurement of the electrical potential from 12 angles, called *leads*



What causes the electrical currents tracked on the ECG?

- The resting state of the heart cells (myocytes) is polarized.
- Excess of potassium (K+) inside, excess of sodium outside (Na+) causes a potential difference.
- Depolarization: Passage of the electric current through the heart muscle (upward wave in ECG)
- Repolarization: All the myocytes return in the resting state (polarized).





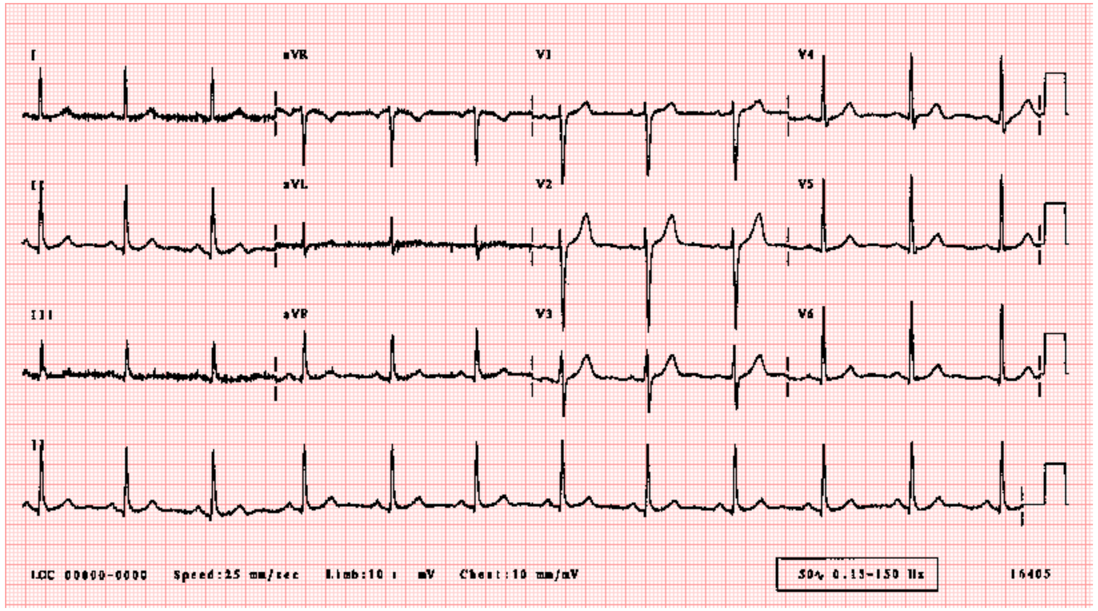
- P wave: atrial depolarization (contraction);
- QRS complex: ventricular depolarization. It hides the relaxation of atria;
- T wave: ventricular repolarization (relaxation)
- U wave: papillary muscle repolarization

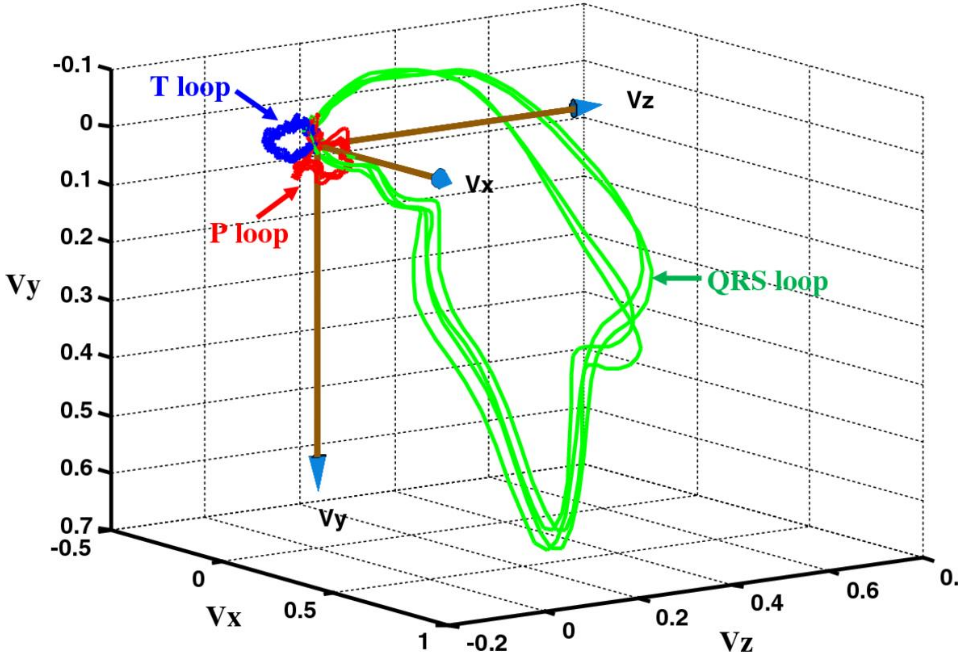
Data

The data were provided by Glostrup Hospital, Copenhagen

- 6667 patients:
  - 1.2 seconds recording of 12-leads ECG (Sampling frequency = 500Hz);
  - Other information, e.g., age at ECG, sex, heart rate (HR), QRS duration, BMI...

Eletrocardiogram (ECG) *Dower's transformation* Vectorcardiogram (VCG)

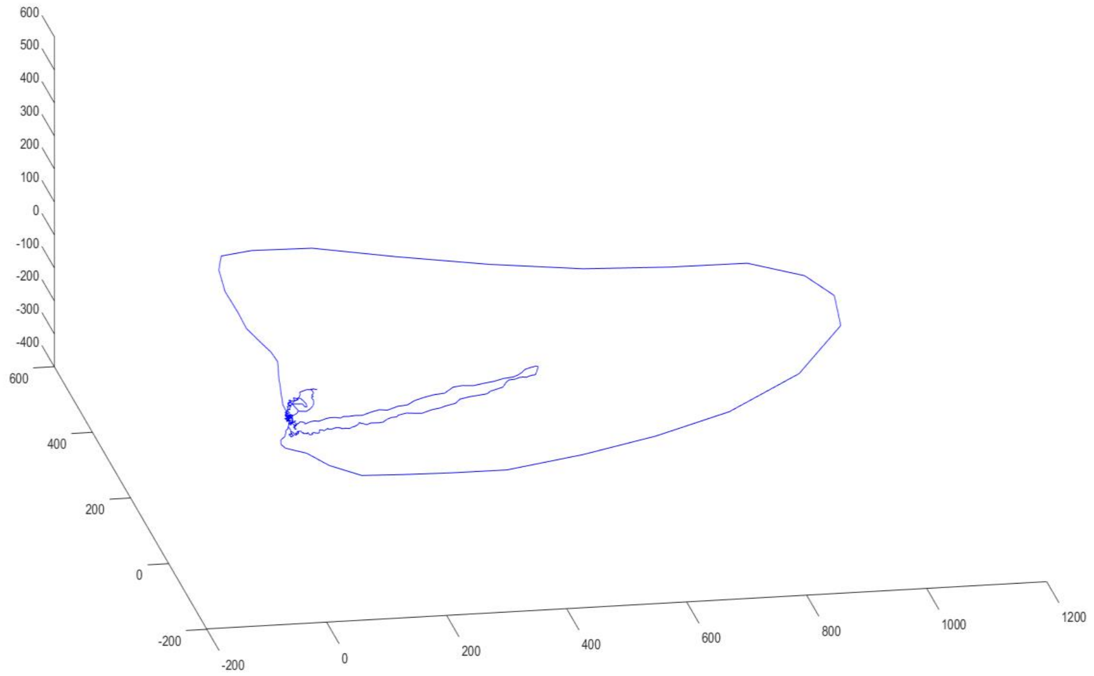


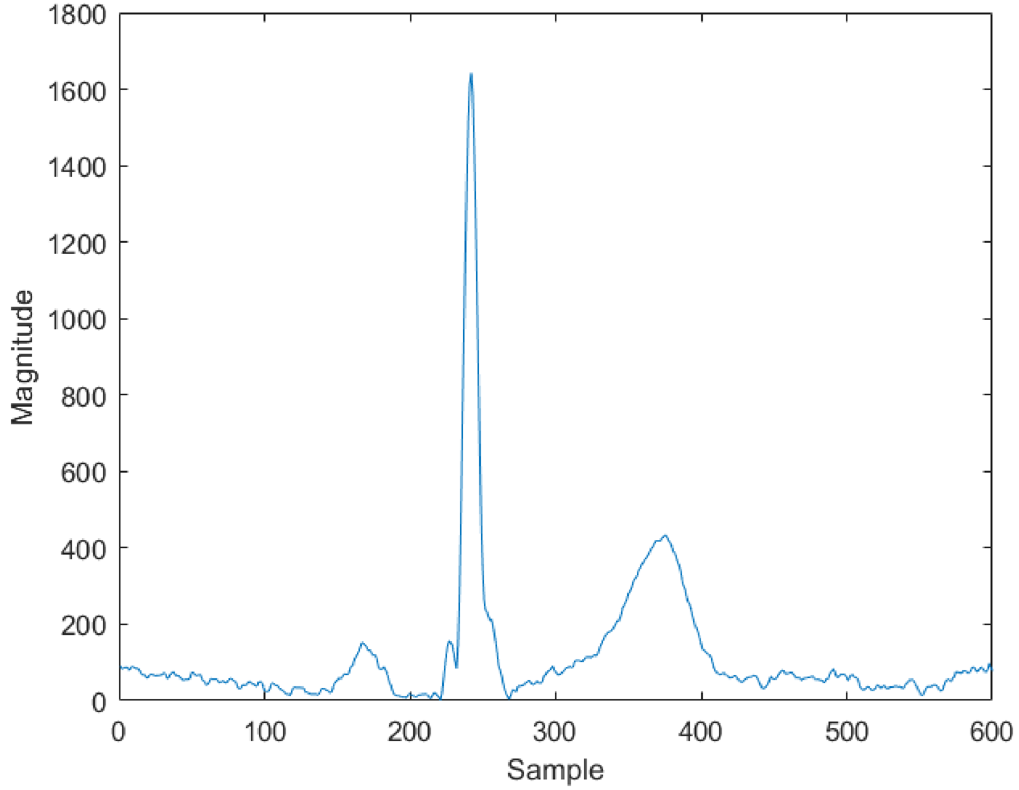


- Recording from 12 angles

- Continuous series of vectors that gives information also about the directions

VCG $\xrightarrow{\quad mag-lead \;=\; ||VCG||_2 \;=\; \sqrt{VCG_x^2 + VCG_y^2 + VCG_z^2}\quad}$ **Magnitude lead**





- Representation of the mangitude of the loops. We lose information about directions.

Clustering

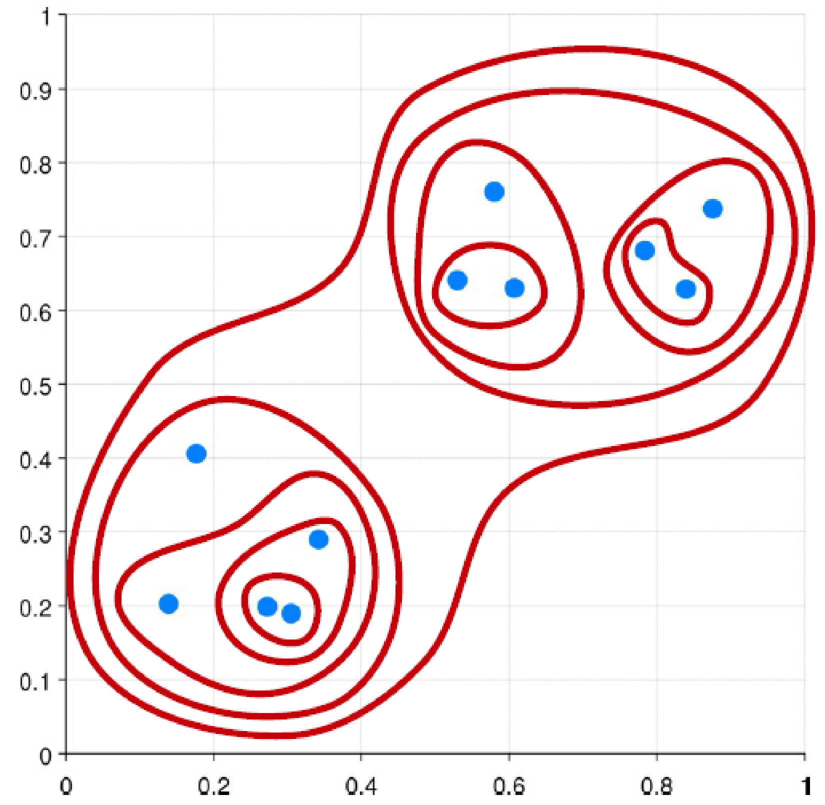
- Unsupervised learning
  - No output variables available, only input data.

- Divide the data into non-overlapping groups which are:
  - **Meaningful**: Capture the natural structure of the data
  - **Useful:** Depends on the purpose

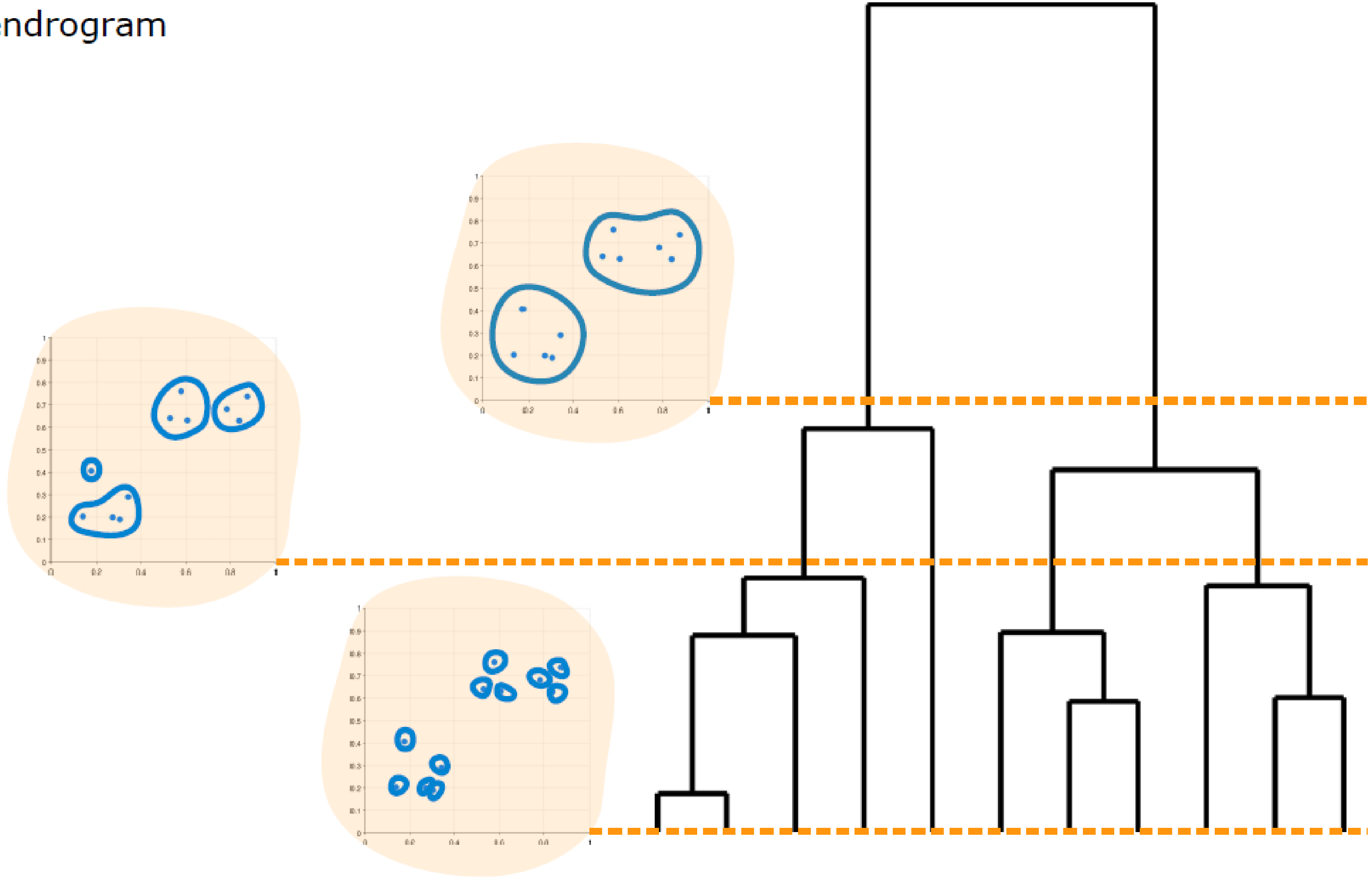
- Observations in the same cluster are **similar in some sense** and observation in different clusters must be different

Agglomerative Hierarchical Clustering

- This is a "bottom up" approach
- Each observation starts in its own cluster
- Pairs of clusters are merged as one moves up the hierarchy.



- Dendrogram



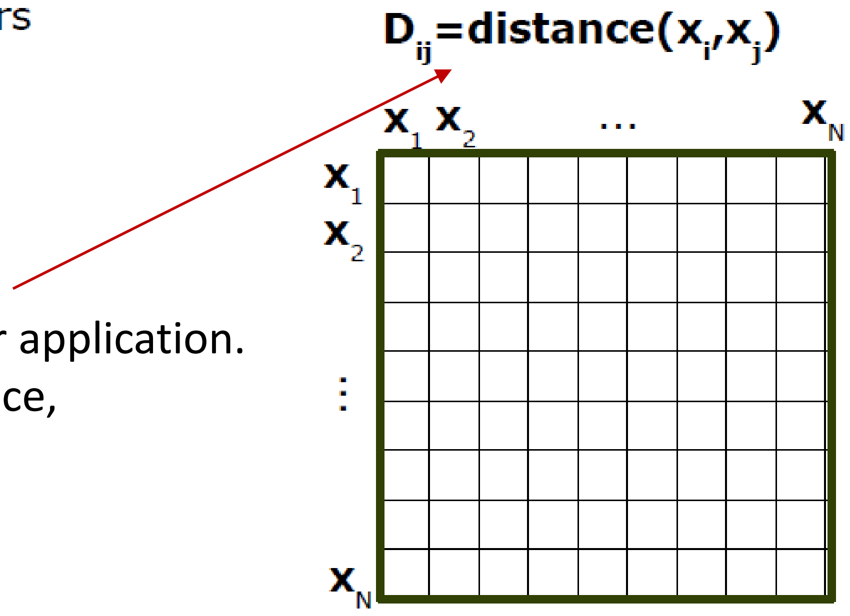
Agglomerative hierarchical clustering

Initialize the proximity matrix

Repeat

- Merge the two closest clusters
- Update the proximity matrix to reflect the proximity between the new cluster and the original clusters

Until only one cluster remains



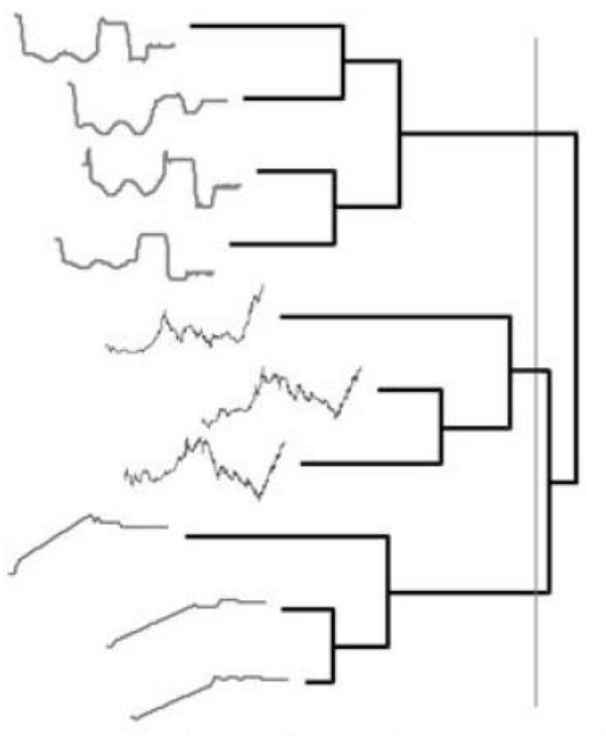
The choice of the distance depends on the particular application. Possible choice are, e.g Euclidean distance, Lp distance, Manhattan distance ecc.

We now perform hierarchical clustering.

<u>Step 1</u>. Build the <u>Proximity matrix</u>

<u>Step2</u>. Build the Dendrogram

<u>Step 3</u>. Cut the dendrogram to the desired level (i.e number of clusters).

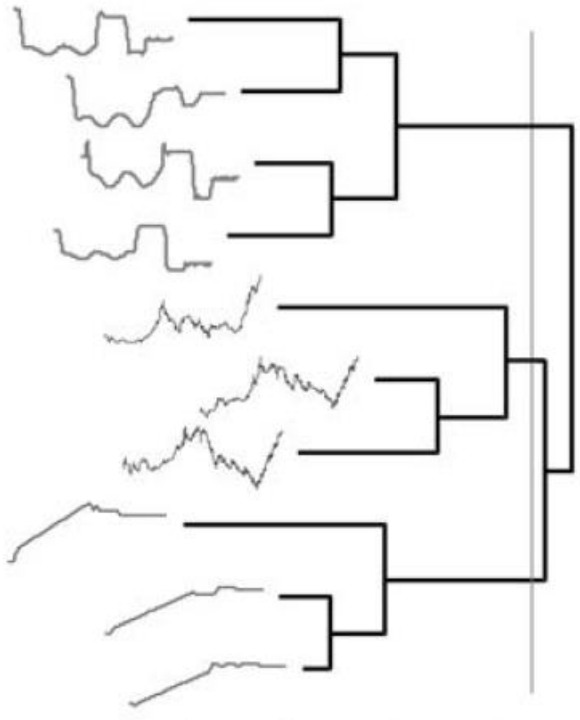


We now perform hierarchical clustering.

Step 1. Build the Proximity matrix

Step2. Build the Dendrogram

Step 3. Cut the dendrogram to the desired level (i.e number of clusters).

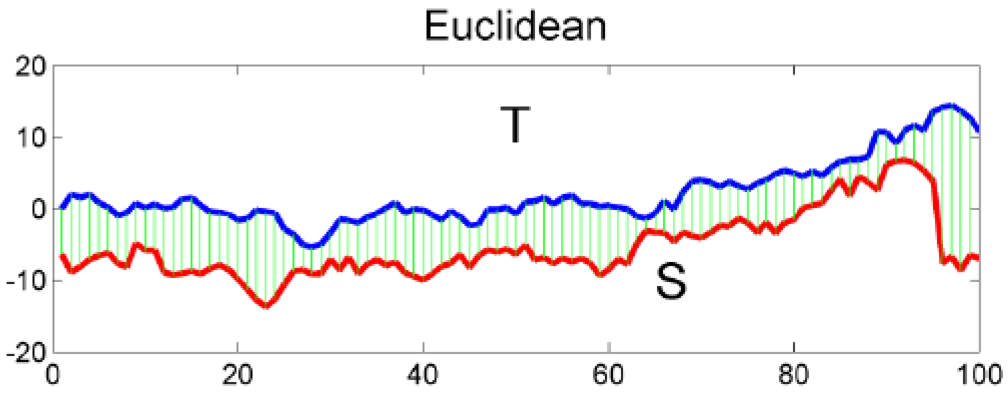


WE NEED THE CONCEPT OF SIMILARITY BETWEEN TIME SERIES

Dissimilarity measure

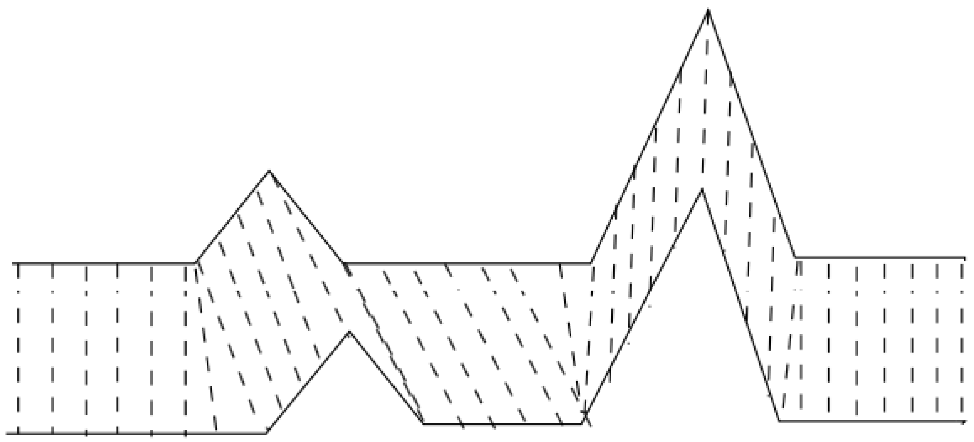
- Our dissimilarity measure for time series must take into account:
  - Intersubject variability
  - Different dutarions of QRS sequences
  - The time correlation between points in the same series

Euclidean distance is not suitable

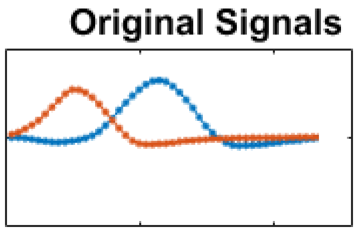


Dynamic Time Warping (DTW)

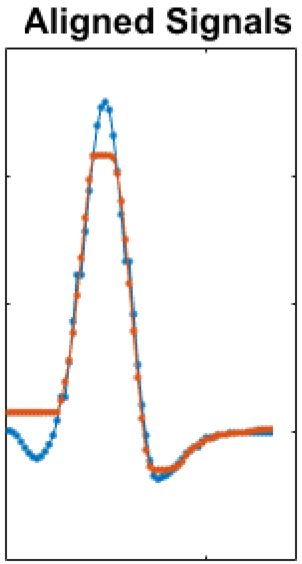
- It is an algorithm for measuring similarity between two temporal sequences, which may vary in speed. DTW computes the optimal match between two time series.



DTW

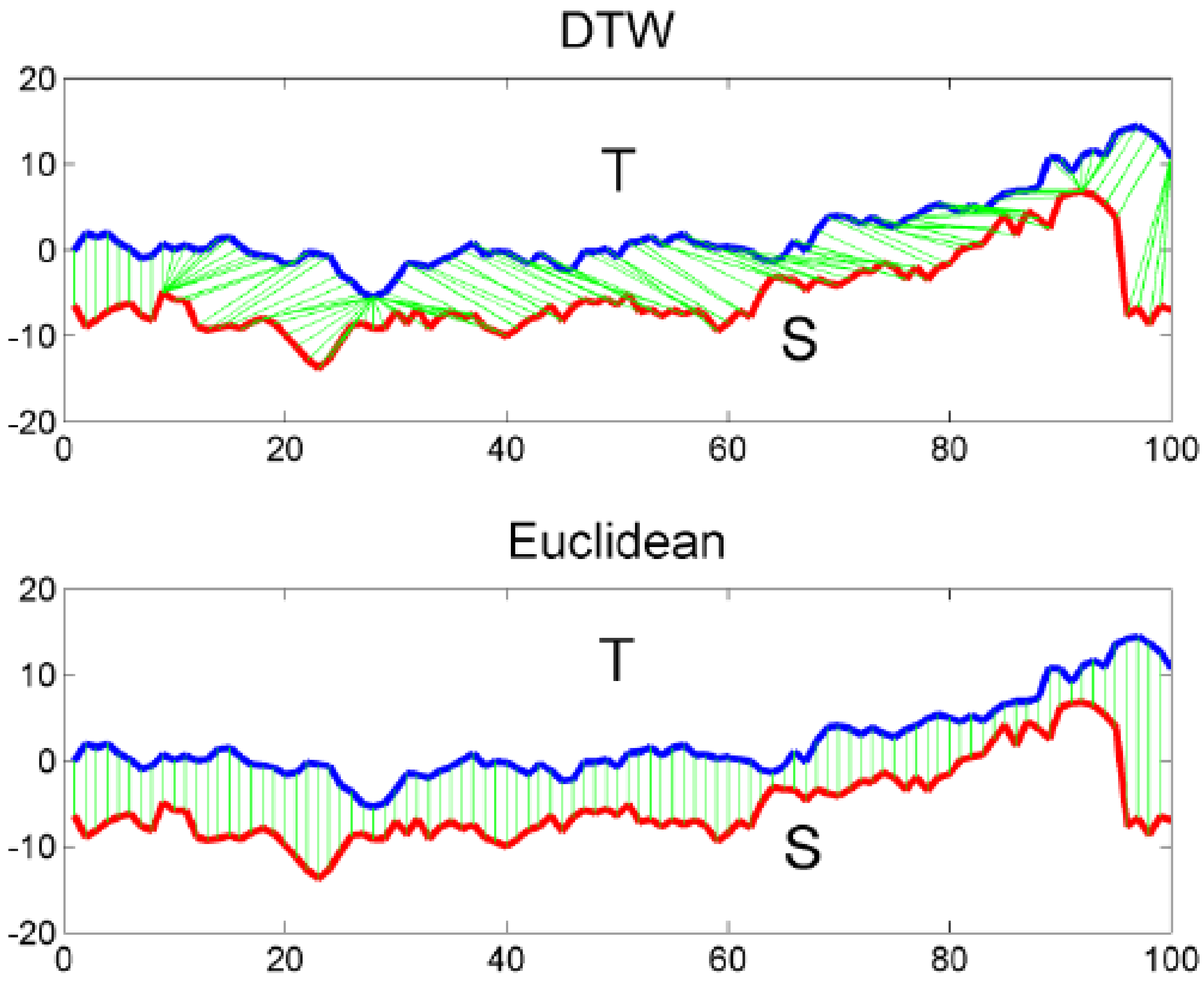


With the DTW algorithm, we can also align the two sequences in order to make them comparable



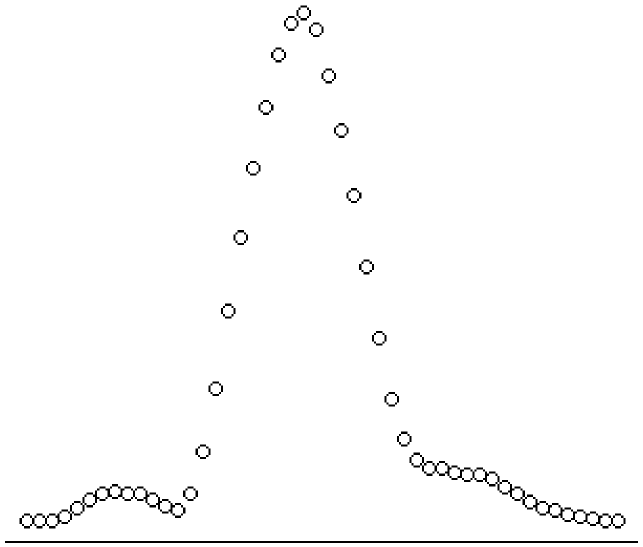
DTW finds the best match between the sequences, time-alignes them and computes the distance between them.

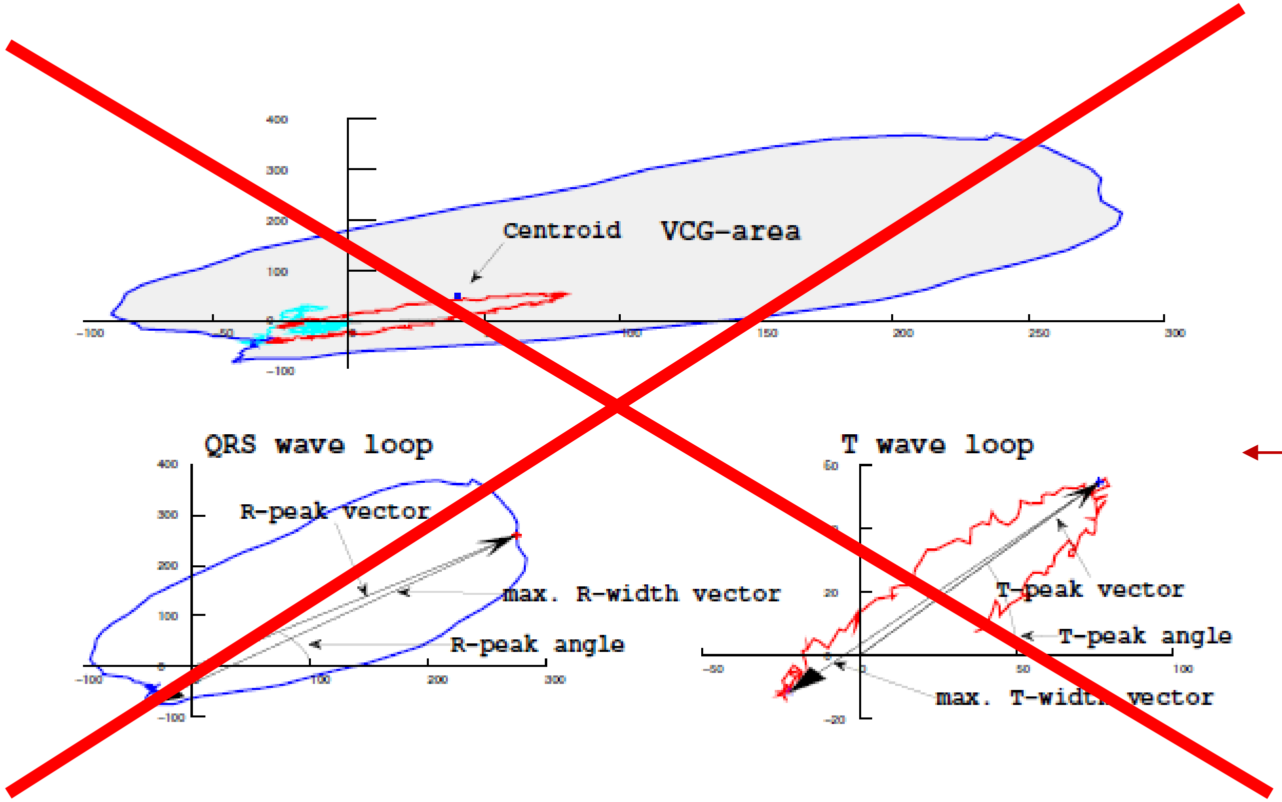
DTW vs. Euclidean distance



Our analysis has been carried out on “raw” QRS loop or T wave, i.e., without feature extraction

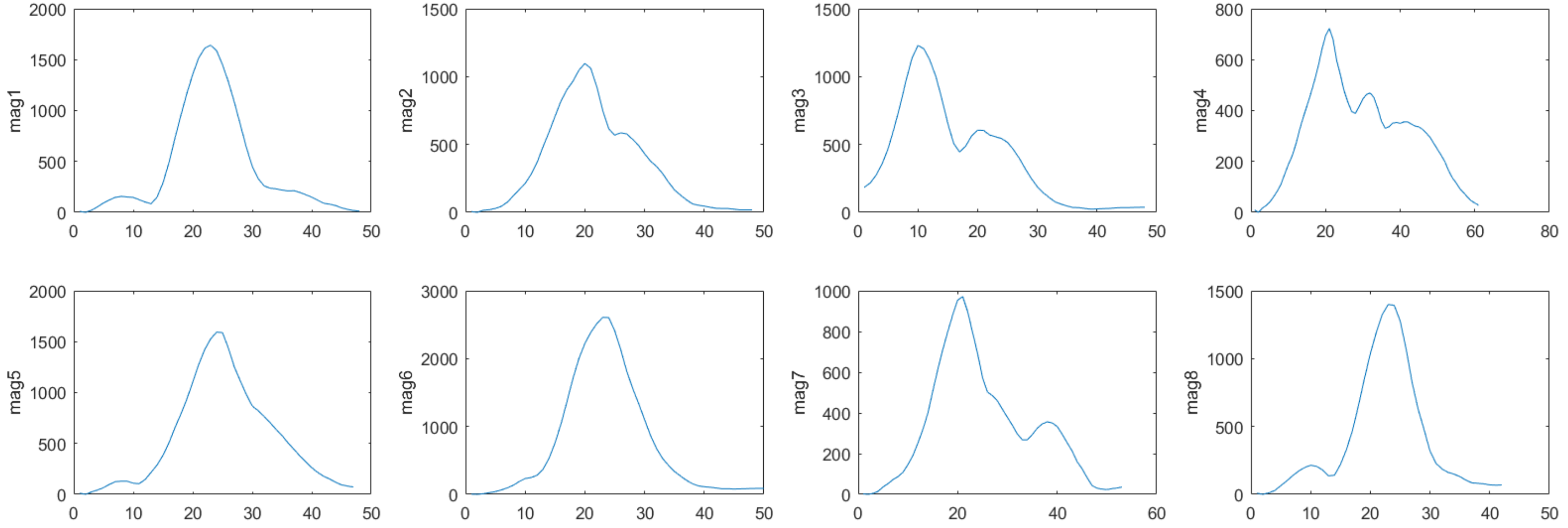
The waves themselves can be used somehow as biomarkers.

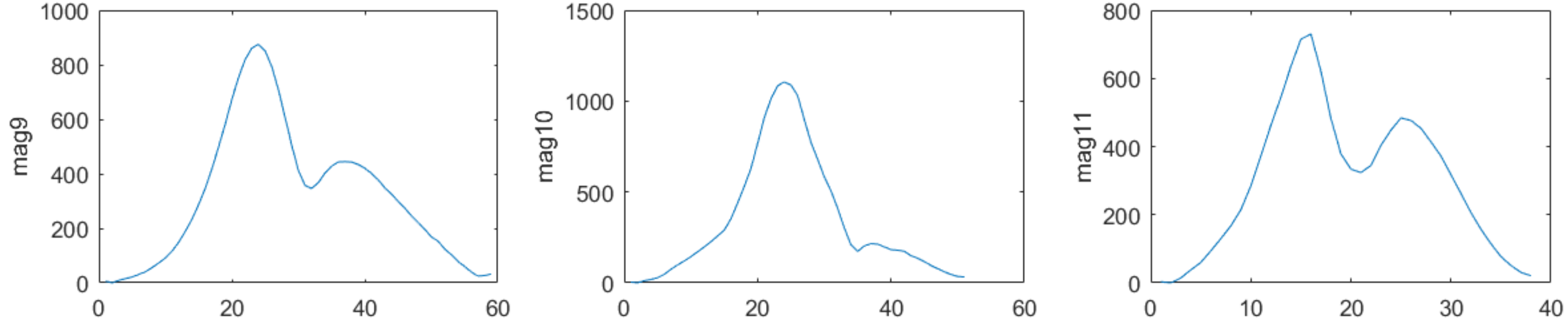




We don’t want to apply the so-called “manual features engineering”.

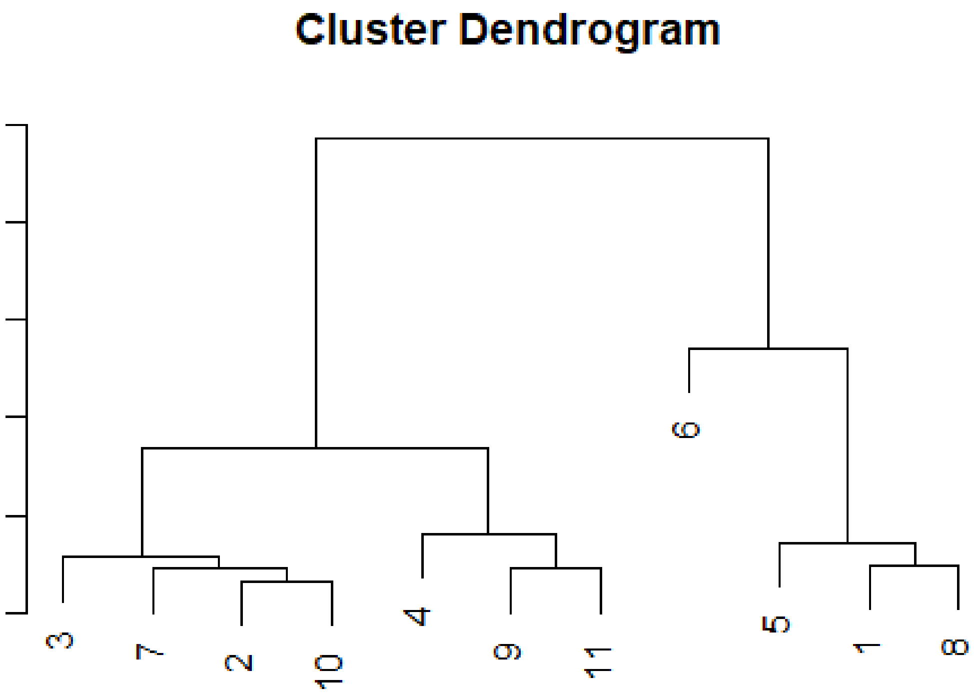
A simple example: 11 patients gropued in 4 clusters

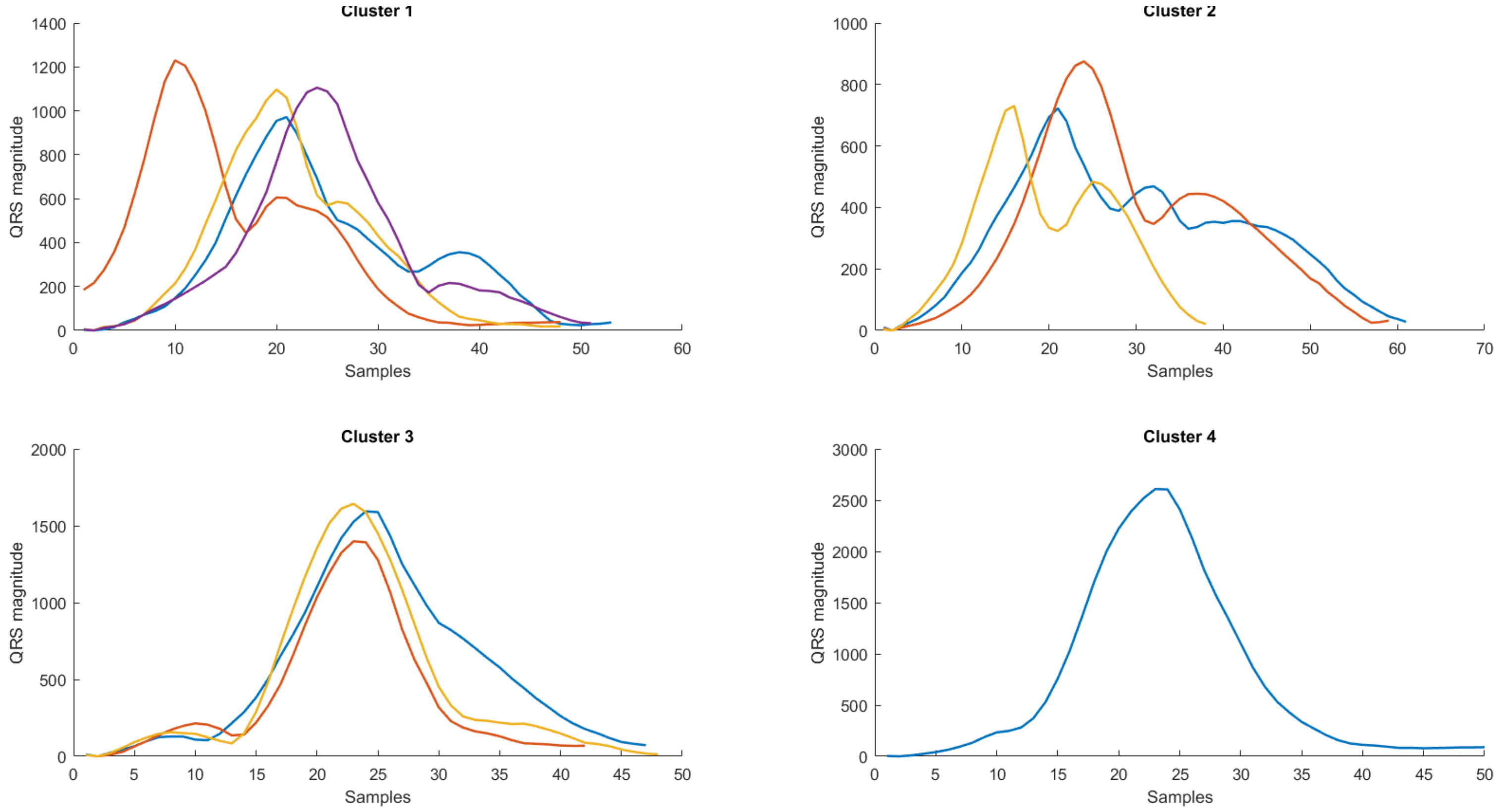




The plots show the magnitude of the QRS loop for each patient

Simple case: only 11 patients grouped in 4 clusters.

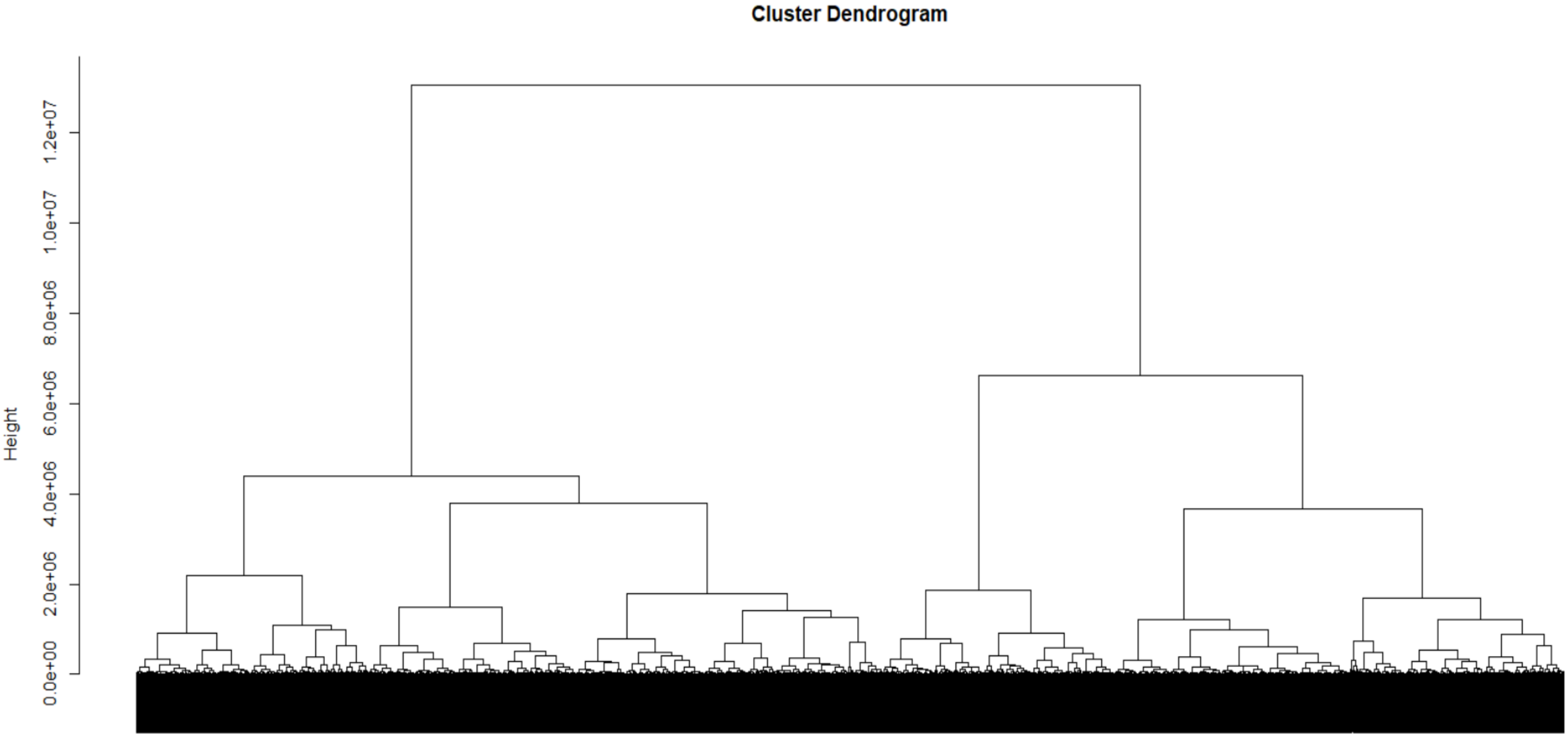




Experiments

- DTW algorithm has been applied to:
    - QRS loop magnitude of VCGs;
    - 3 dimensional QRS loop of VCGs;
    - 3 dimensional T loop of VCGs;
    - T-wave of the 12 leads ECG.

For all the experiments, $k = \{4, 6, 10\}$.


Cluster Dendrogram
Height
0.0e+00
2.0e+06
4.0e+06
6.0e+06
8.0e+06
1.0e+07
1.2e+07

???

Results – Focus on gender

Cluster n.	1	2	3	4	p-value
vcg_mag_qrs	56.3	43.8	41.2	73.0	<0.001
vcg_qrs	40.6	52.0	44.8	61.7	<0.001
vcg_tloop	**18.7**	51.4	**85.6**	63.2	<0.001
ecg_twave	43.4	76.6	**19.0**	51.7	<0.001

Cluster n.	1	2	3	4	5	6	p-value
vcg_mag_qrs	58.9	53.5	43.8	42.9	38.6	73.0	<0.001
vcg_qrs	66.0	**17.7**	43.7	64.4	44.8	61.7	<0.001
vcg_tloop	**19.7**	43.5	**17.0**	**85.6**	63.2	70.4	<0.001
ecg_twave	38.0	51.1	**82.6**	**19.0**	70.8	51.7	<0.001

Cluster n.	1	2	3	4	5	6	7	8	9	10	p-value
vcg_mag_qrs	54.7	63.0	53.5	40.0	42.9	47.8	38.9	37.7	72.0	**80.5**	<0.001
vcg_qrs	45.0	73.3	**17.7**	48.0	64.4	37.7	23.2	60.1	63.0	65.8	<0.001
vcg_tloop	21.7	46.8	**17.0**	38.1	**16.6**	**86.3**	**84.7**	72.0	55.0	70.4	<0.001
ecg_twave	32.7	71.9	27.3	**82.6**	61.9	**15.9**	58.0	**20.7**	70.8	51.7	<0.001

Future work

- Bulid a similarity measure (different from DTW) capable to perform better on measuring how two ECGs or 3D VCGs are similar;
- Use QRS loop (or other loops) as biomarkers for some heart diseases;
- Finally carry out classification and predict the risk of an individual to be prone to heart disease.

Thank you!