

---

# The Patterns of Conflict Emergence (PaCE)

Thomas Chadeaux  
Trinity College Dublin

15 July 2022

---

# PaCE

---

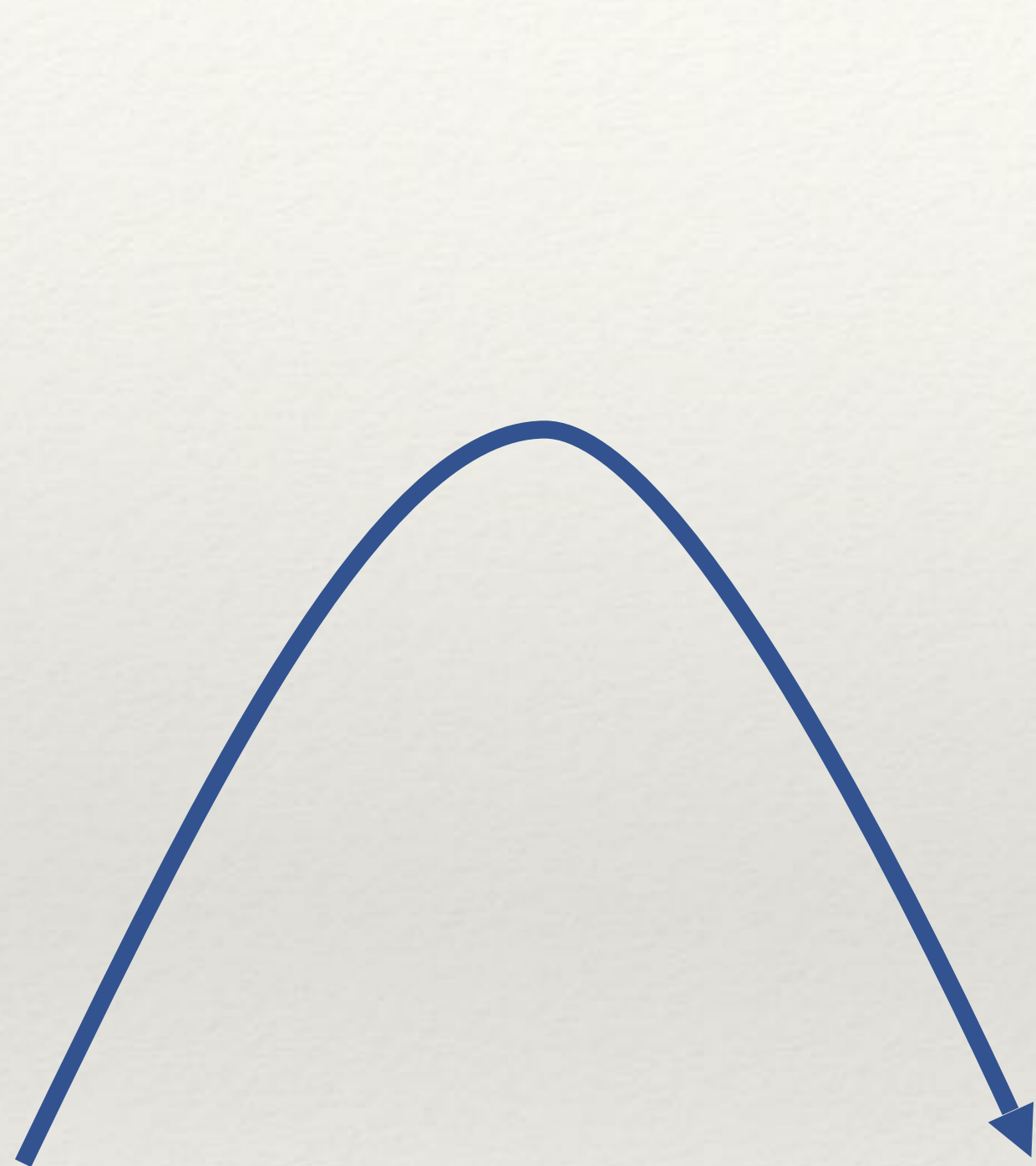
- ❖ **ERC 2022-27**
- ❖ **Trinity College Dublin, Department of Political Science**
- ❖ **Team:**
  - ❖ **Political Science**
  - ❖ **Economics**
  - ❖ **Physics**
  - ❖ **Machine Learning**

# PaCE: Project overview

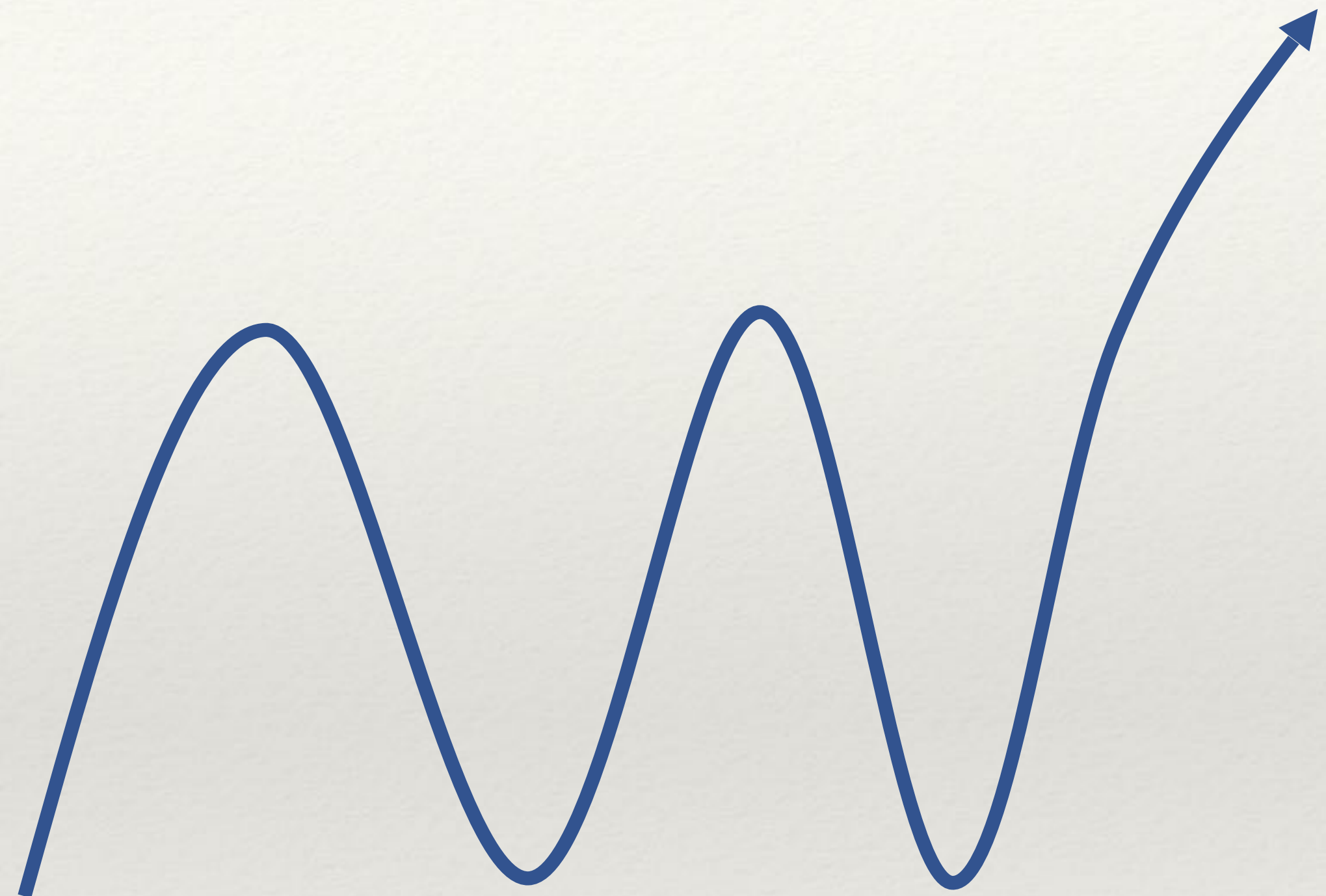
---

# Stories

---

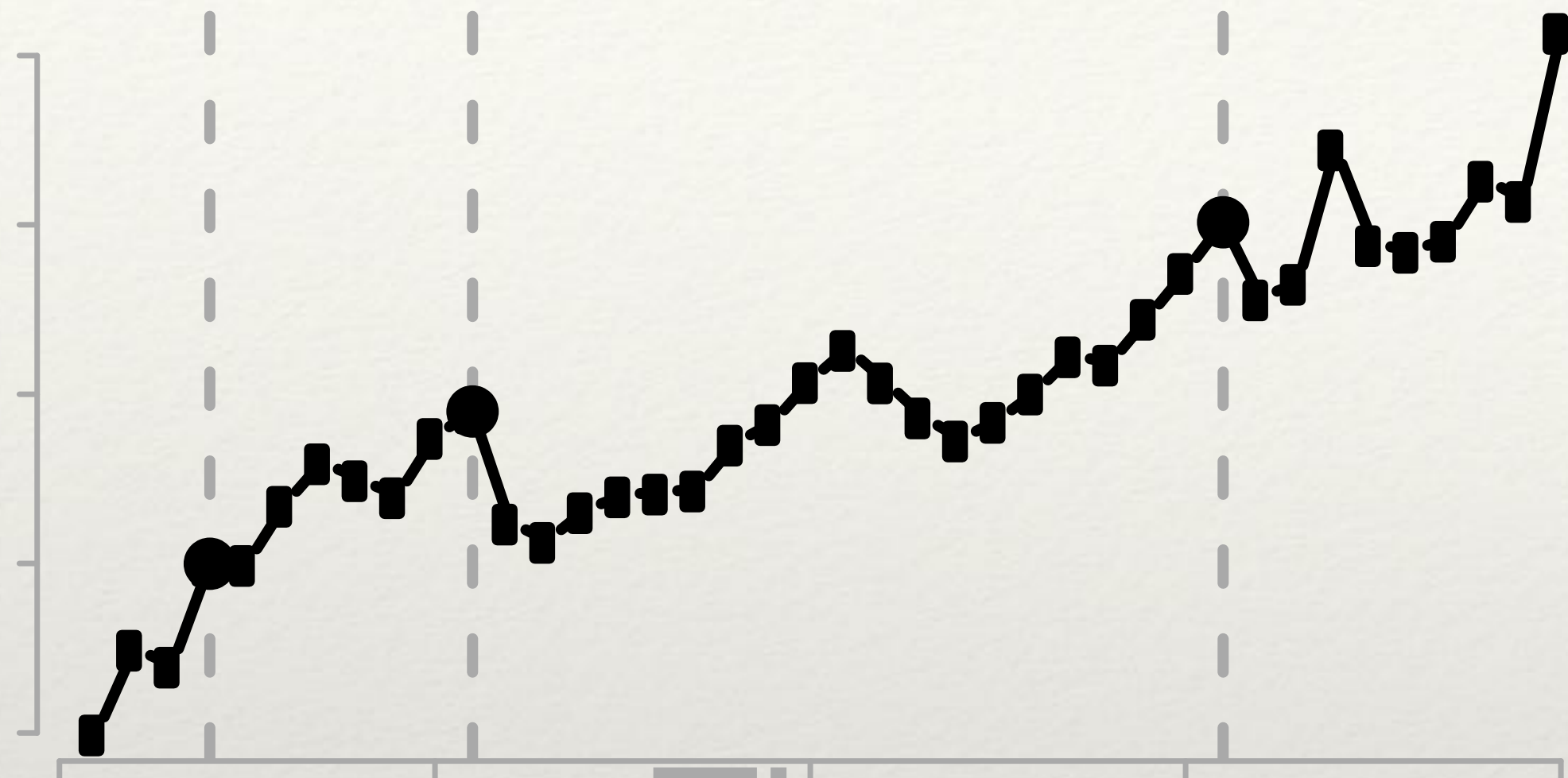


Tragedy



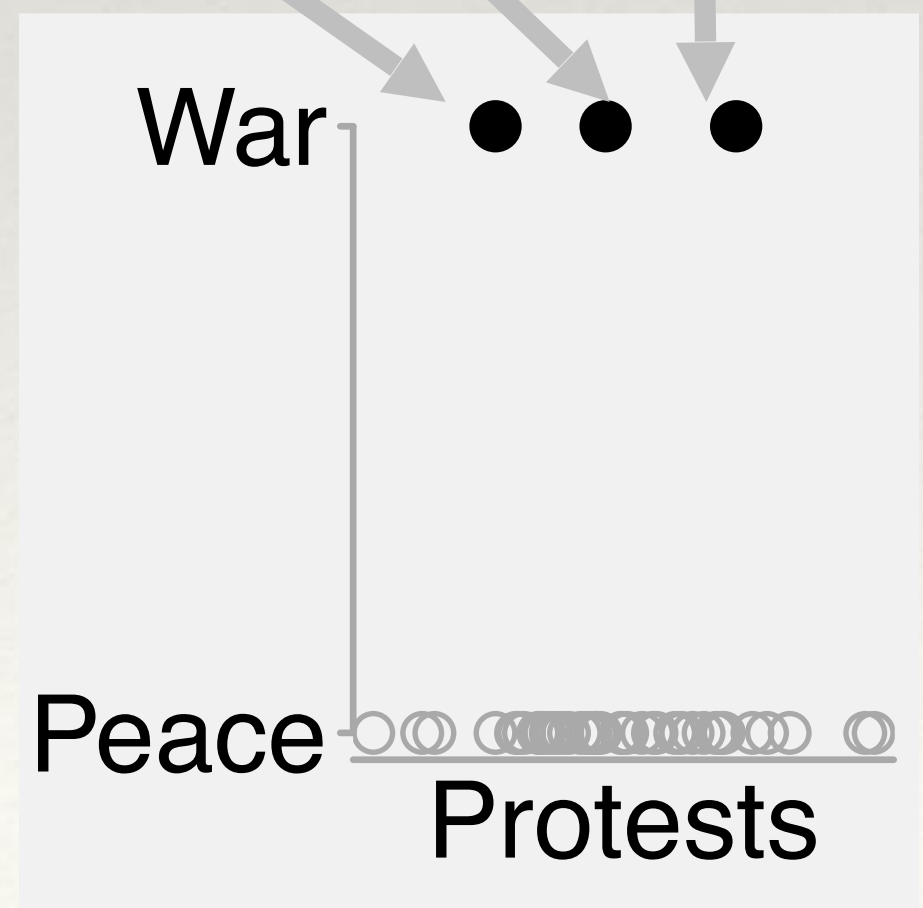
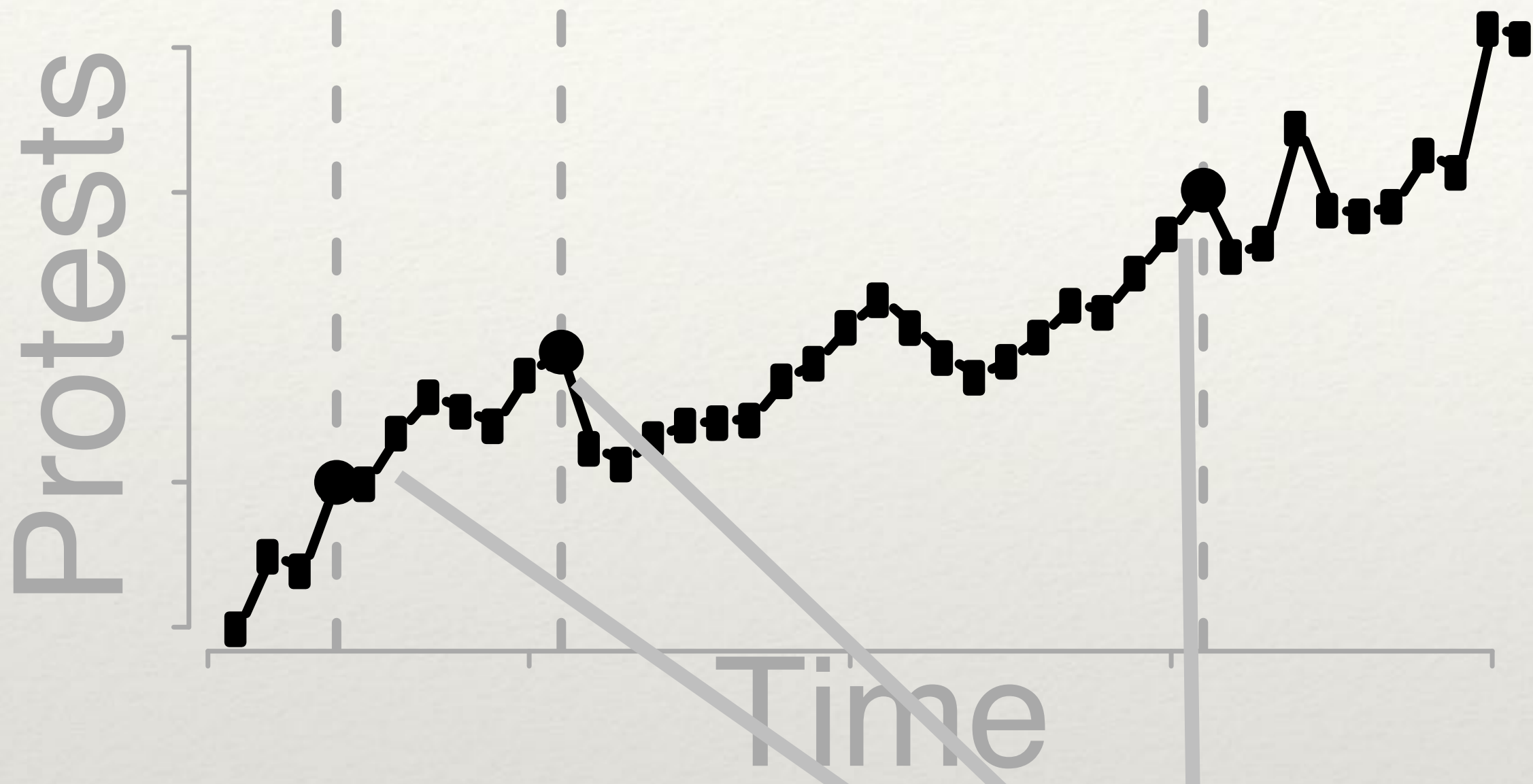
Seven point story

Protests

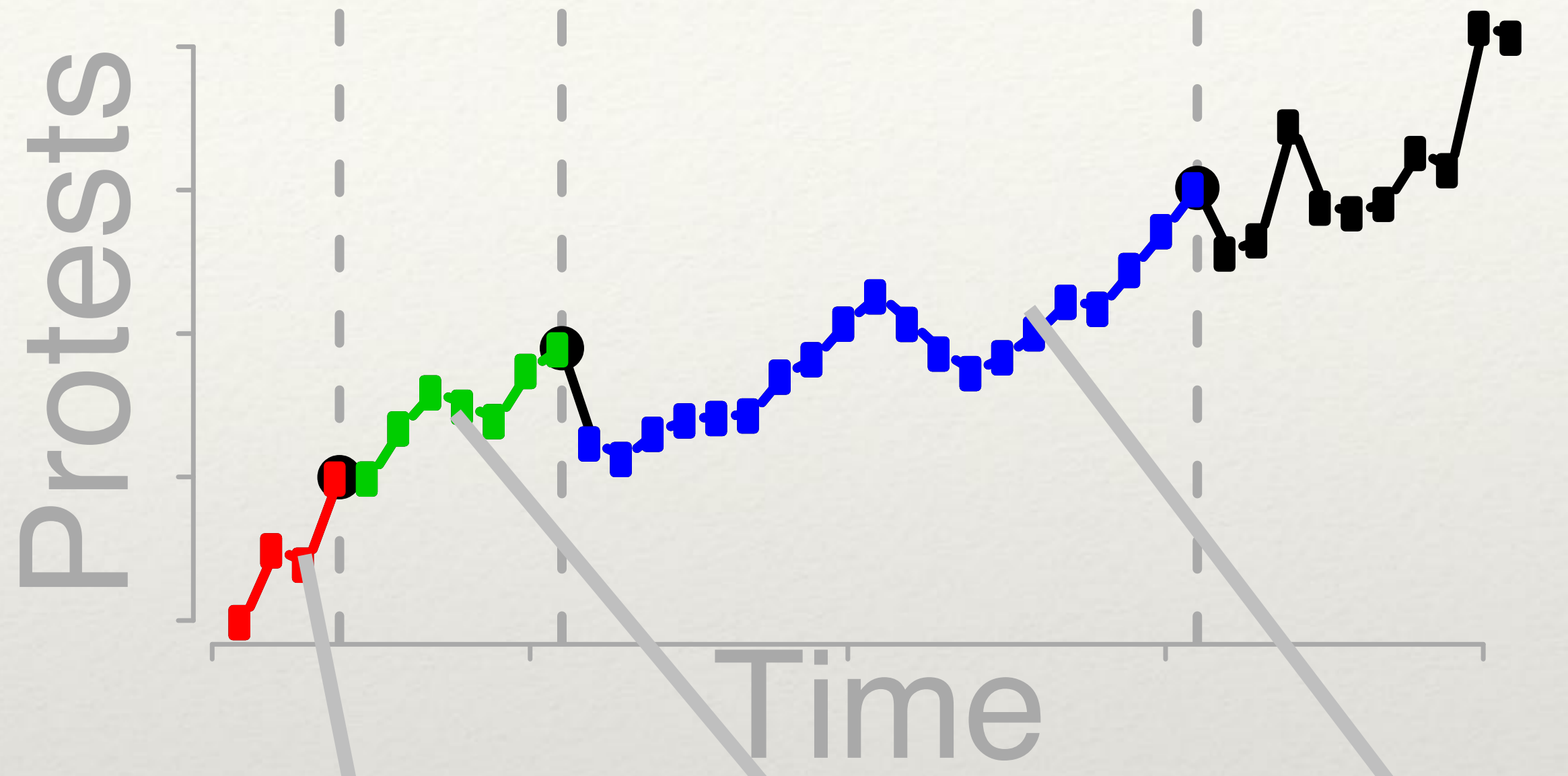


Time

# Observations- as units of analysis



# *Sequences* as units of analysis



# Sequences in other fields

```
12854400 tcaaagtaagttagataaacatgatcattcacagggtcagatggttttaaaaaaaaaatcattatgggtgtacatcacatgtagacaataacttcagaattcatc
12854200 taggaaaagttaatggttacggccaatcacttttttaacagcccaaacacataatagctccaaatatcattttttcccctagaatattctcaacct
12854000 cttgtaaagtattcacatttcattccaagaaaaatagactgatgaagaaatataatcagatatgacaaggccgtgtcgttttaggttacgtaactctaca
12853800 aggtttagggtctcaatataaacacacaaaagcagatagaagaagcaaaccattcacaatcagacaATGACATCTCTCCATACGTTACTCTTCTCTTCTCT
12853600 TCTTTTCTTCATCGTCTTCCAACCTTCACGTTTTCTCCACCTTATTGTTTCAGgttcgctcttttagttttgcttctttacatacacagactctacacac
12853400 tcacttattgggtttctttcaattgtgaaacagAGTTTCAATTGGGAGTCATGGAAGAAAGAAGGAGGATTCTACAATTCTCTCCACAACCTCCATTGACG
12853200 ACATAGCCAACGCTGGAATCACTCATCTTTGGCTTCCTCCTCCTTCTCAATCCGTTGCTCCTGAAGgttcatttctgctttactctttacacattcaca
12853000 taccaatcttggttactcacgcaatcttcattcctcagGTTACTTACCGGGAAAGCTATACGATCTAAACAGCTCCAAATACGGTTCAGAGGCGGAACTGA
12852800 AATCGTTAATCAAAGCGTTGAATCAAAAAGGAATAAAAGCTTTGGCTGATATAGTGATTAACCACAGAACAGCTGAGAGGAAAGACGATAAATGTGGATA
12852600 CTGTTATTTGGAAGGTGGGACTTCCGATGATCGTCTTGATTGGGATCCTTCCTTTGTCTGCCGCAATGACCCTAAATTTCCCGGTACCGGAAACCTCGAC
12852400 ACCGGAGGAGATTTTGATGGAGCGCCCGACATCGACCACCTTAACCCTAGAGTTCAGAAAGAGTTGTCCGAATGGATGAATTGGCTTAAAACCTGAAATCG
12852200 GATTCCATGGTTGGAGATTTGATTATGTTTCGAGGTTATGCATCTTCCATCACCAAATTATACGTTTCAGgtaaatacacatatgaattctcaaatatcagac
12852000 aacagtattagtatataagaaacatagggttgagataattatttactattagttatataagttatcatagggttgatagggttatttactactatttagt
12851800 ataagaaacataagtcaatgcaatcaataagaaatataagaaagttcactactgattatgtgataaattcctctgtttttggatacacagaAATACATC
12851600 ACCGGATTTTGCGGTGGGTGAGAAATGGGACGATATGAAGTACGGAGGAGACGGGAAACTAGACTATGATCAGAACGAGCATCGGTCCGGTCTCAAACAG
12851400 TGGATCGAGGAAGCGGGTGGTGGTGTGTTGACAGCTTTTGATTTACACCACAAAGGGATCTTACAGTCTGCTGTCAAAGGTGAGCTTTGGAGACTAAAGG
12851200 ACTCGCAGGGAAAACCGCCTGGTATGATAGGAATCATGCCCGGAAACGCTGTCACATTCATAGATAACCATGATACATTCAGAACGTGGGTTTTCCCTTC
12851000 TGATAAAGTCTTGCTTGGATACGTTTATATACTTACTCATCCAGGAACTCCTTGCATTgtaagtatcatttttagttagttagctataactatttacaactac
12850800 aatcttggtgatatgttatttttggtgcagTTTTATAATCATTACATAGAATGGGGACTAAAAGAGAGCATCTCAAAGCTGGTGGCTATCAGGAACAAAA
12850600 ATGGGATTGGTAGCACAAGCTCTGTAACGATAAAAAGCGGCAGAGGCGGATCTCTACTTGGCTATGATTGATGATAAAGTTATCATGAAGATTGGACCAA
12850400 GCAAGATGTGGGAACACTTGTTCCTTCTAATTTTGCTTTAGCTTATTCAGGCCTTGACTTTGCTGTCTGGGAGAAGAAGTAAcgcataactcgaatcata
12850200 agaaaagtaatcgaatgtatcttcttcttttaataaaaacattttggcagtatctaaagatatgtataatgaaatataaaatgataaagaatacctaaa
12850000 taaaaagagcactagtggtggtaaaggatacaactccagtgaagaaagaggtcaagtgaagaagtgtcaactgttagaaataagtattggaaagtttc
12849800 catcgttttggtttggtgcatacaactaatatattatattggccgactcgtataagatttgagccctactaaaatcagaattatgatgtcttaacca
12849600 cacaatactgccaaatcagaacgaattatatttgtagaagaagaaaaaaaagtatgggtgggaagtggaacagttagacaggtaaatcgaataaa
```

A  
T  
4  
G  
2  
5  
0  
0  
0  
.1

v

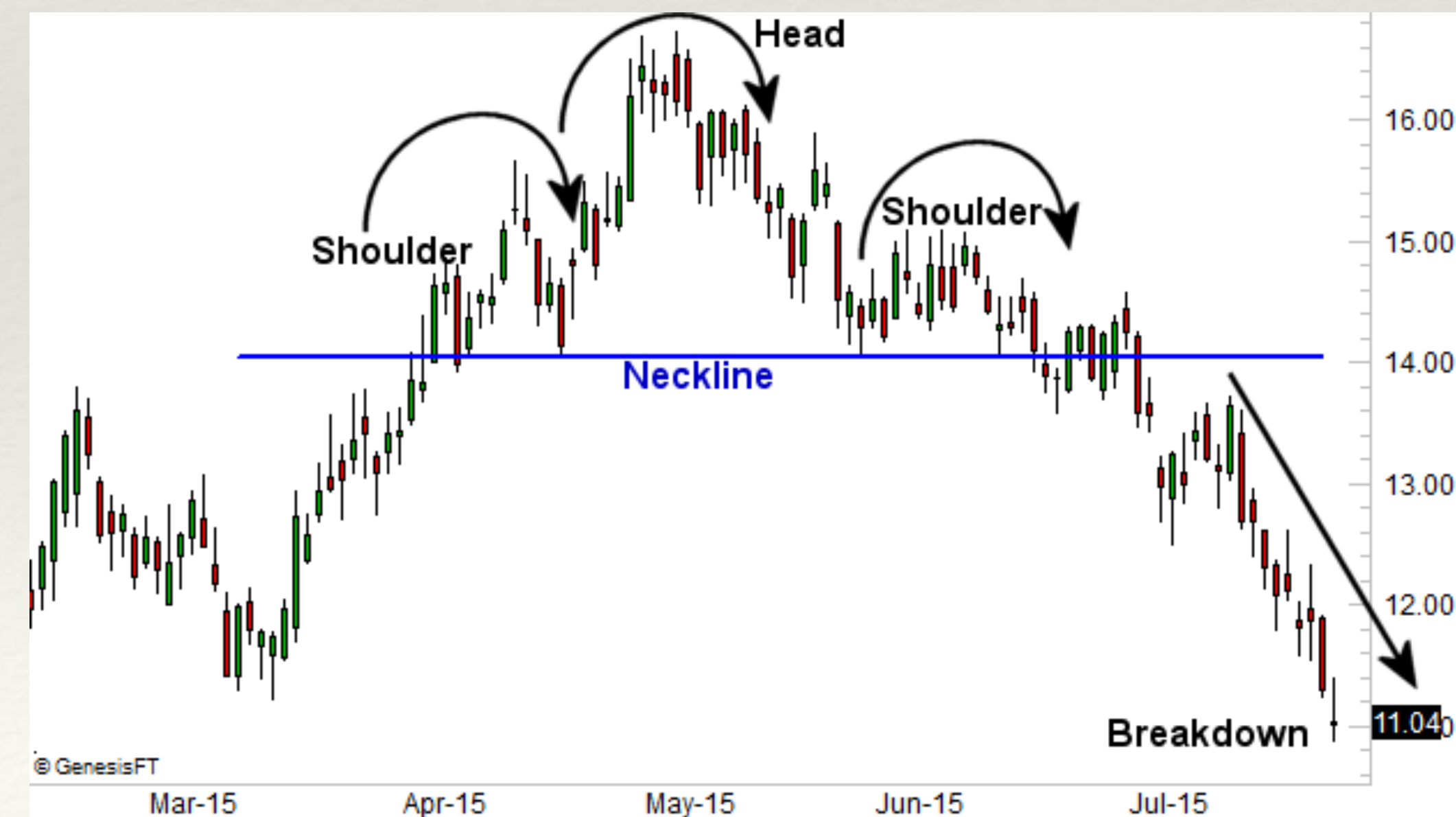
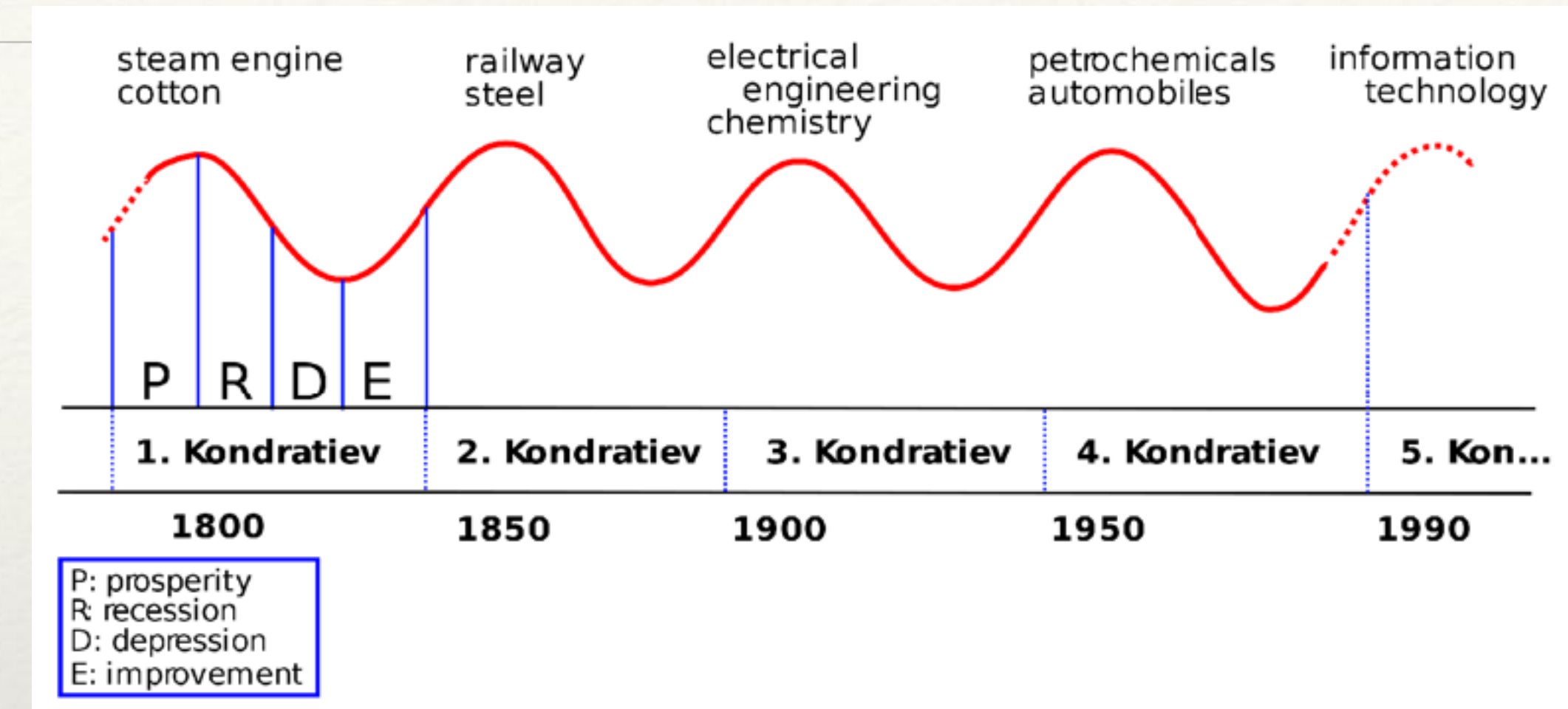
# In social sciences

- ❖ Patterns

- ❖ Marx: Repeating motif over time
- ❖ Kondratiev waves: alternating intervals of high sectoral growth and intervals of relatively slow growth

- ❖ No patterns

- ❖ Efficient market hypothesis
- ❖ Richardson





---

# PaCE: One Big Question

---

Does history repeat itself?

---

# PaCE: Three questions

---

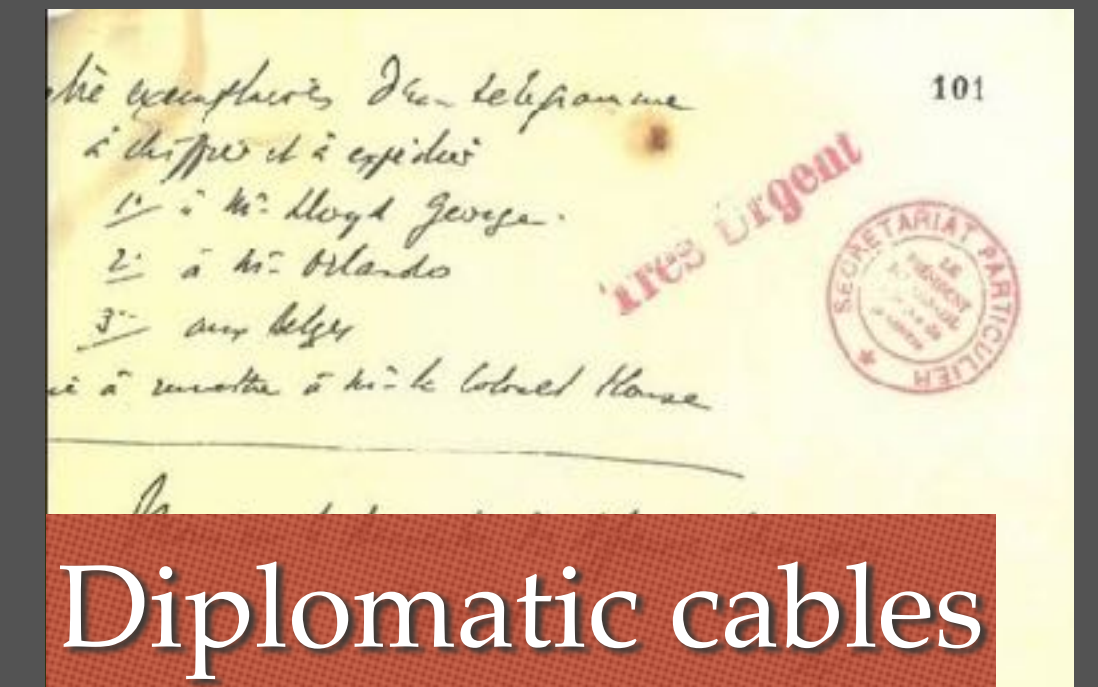
1. Are there patterns in the temporal sequences that lead to conflict?
2. Can we use these motifs to predict conflict?
3. Can we cluster these patterns to create new / inform existing theories of conflict

How?

# Predict...

- Interstate and Civil wars / conflict
- Migration flows
- Protests
- Violence
- and more

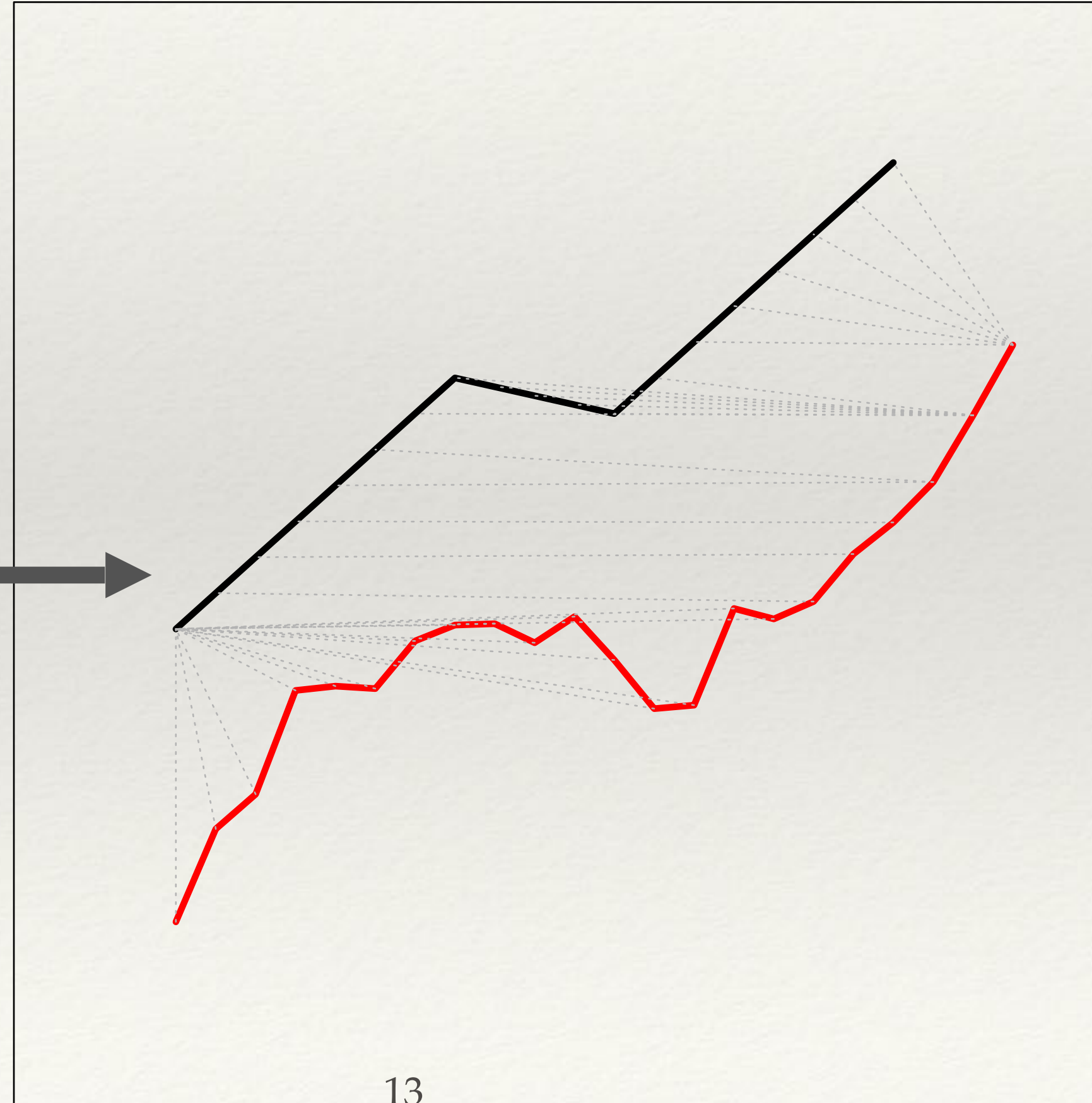
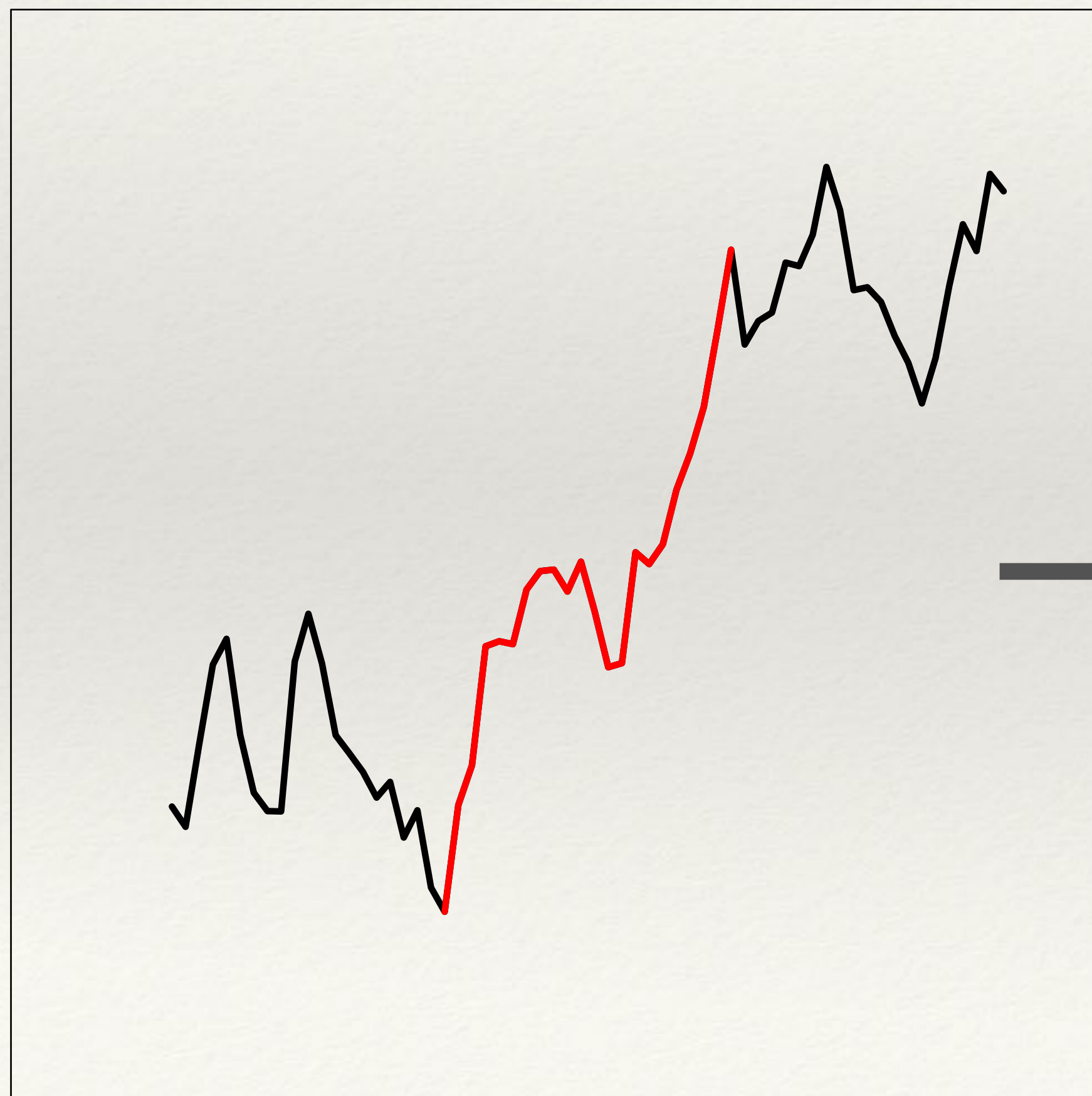
... using



Data acquisition  
& processing

Distance  
measures

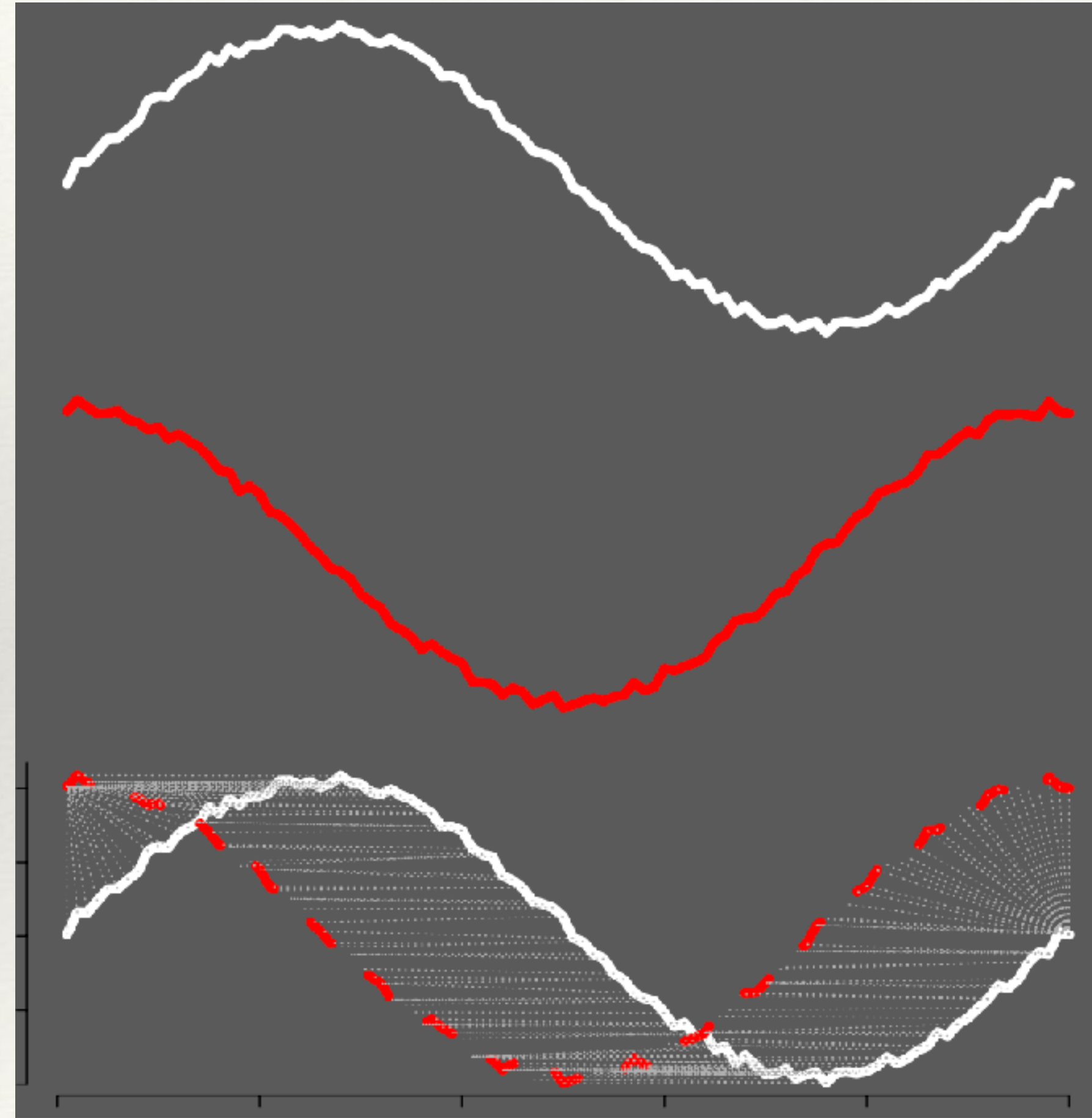
Classification



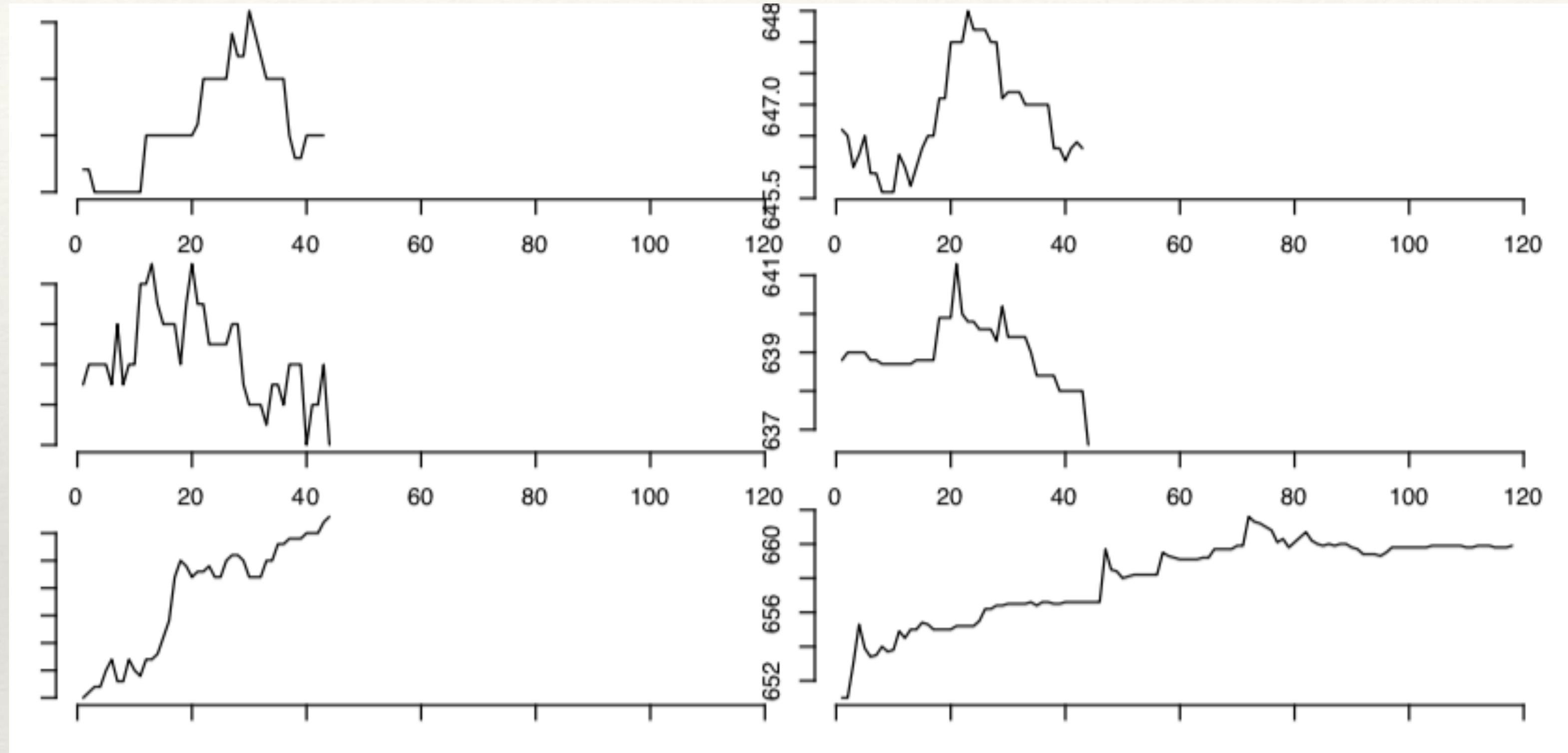
---

# Distance measures: An example

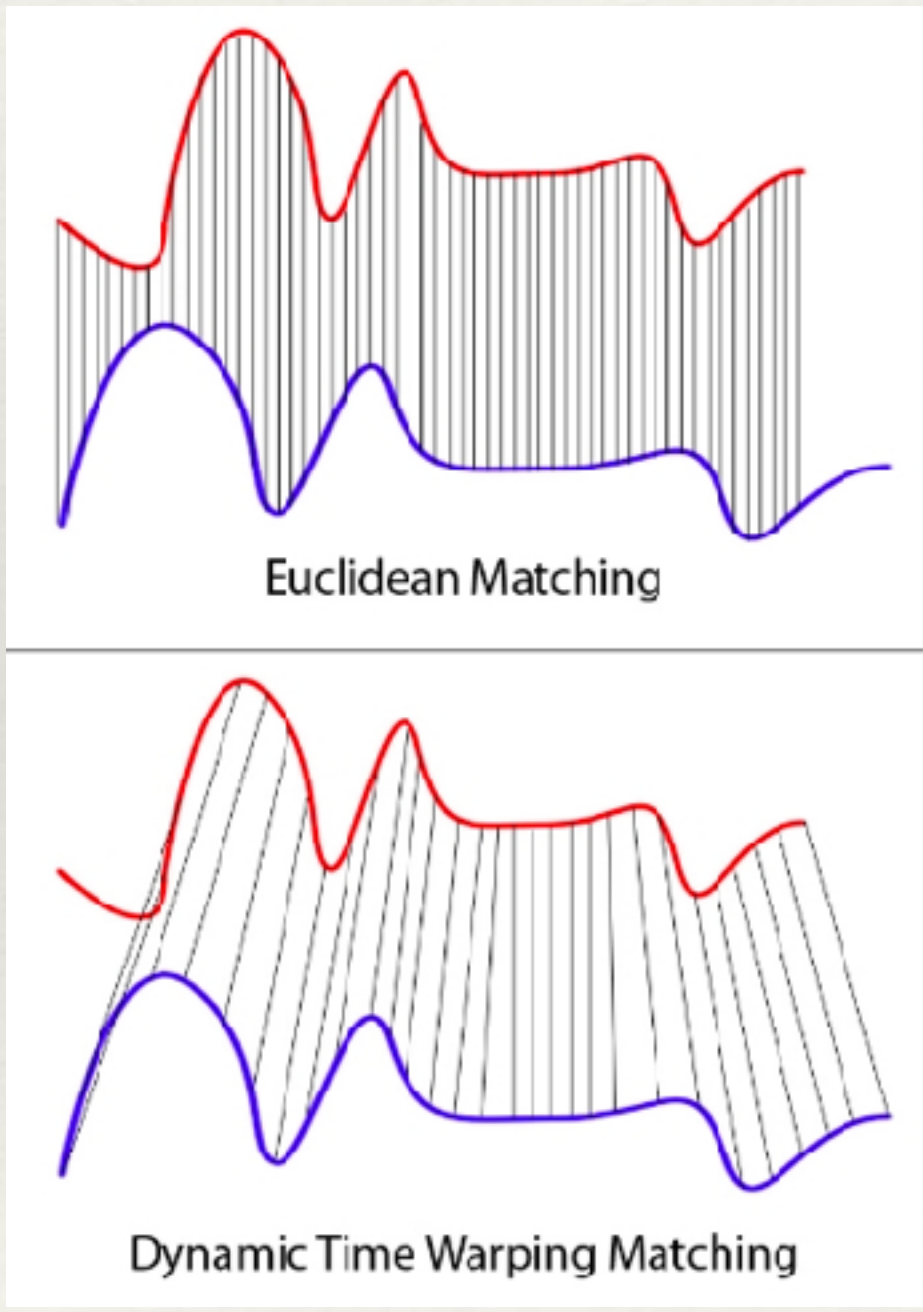
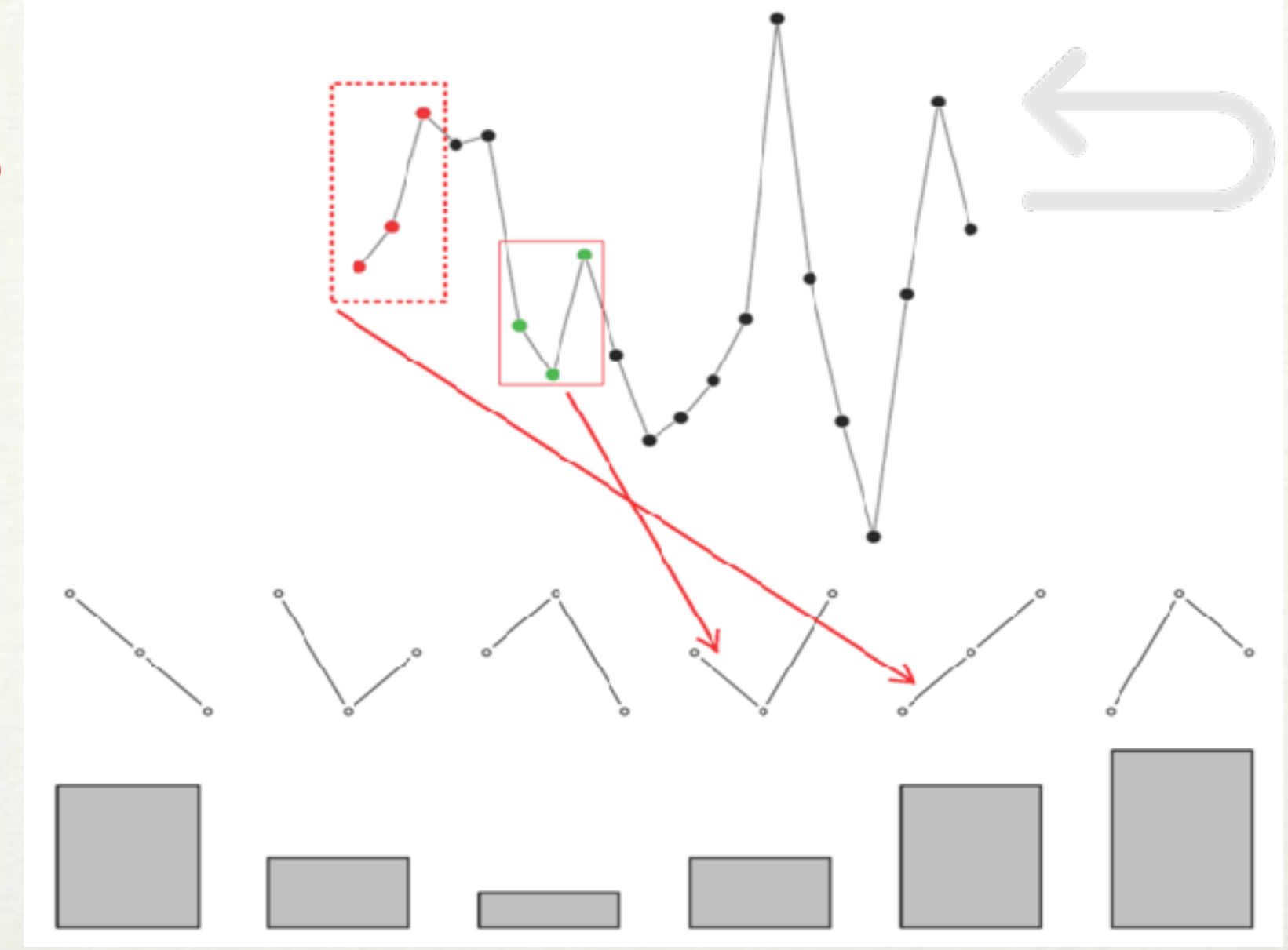
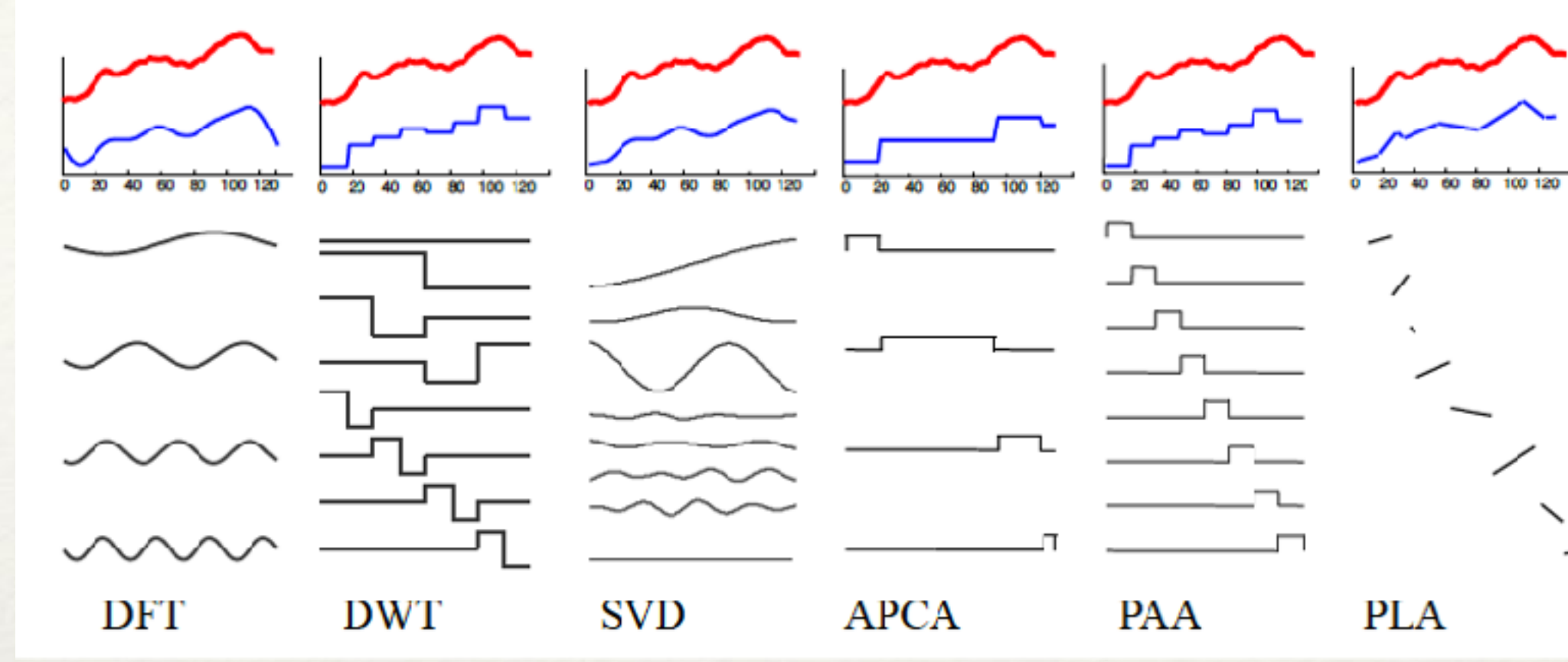
---



Dynamic Time Warping



# Methodology: Distance Measures



Dimensionality Reduction

Permutation Distribution

Clustering

Clustering

Predictions

Dynamic Time Warping



---

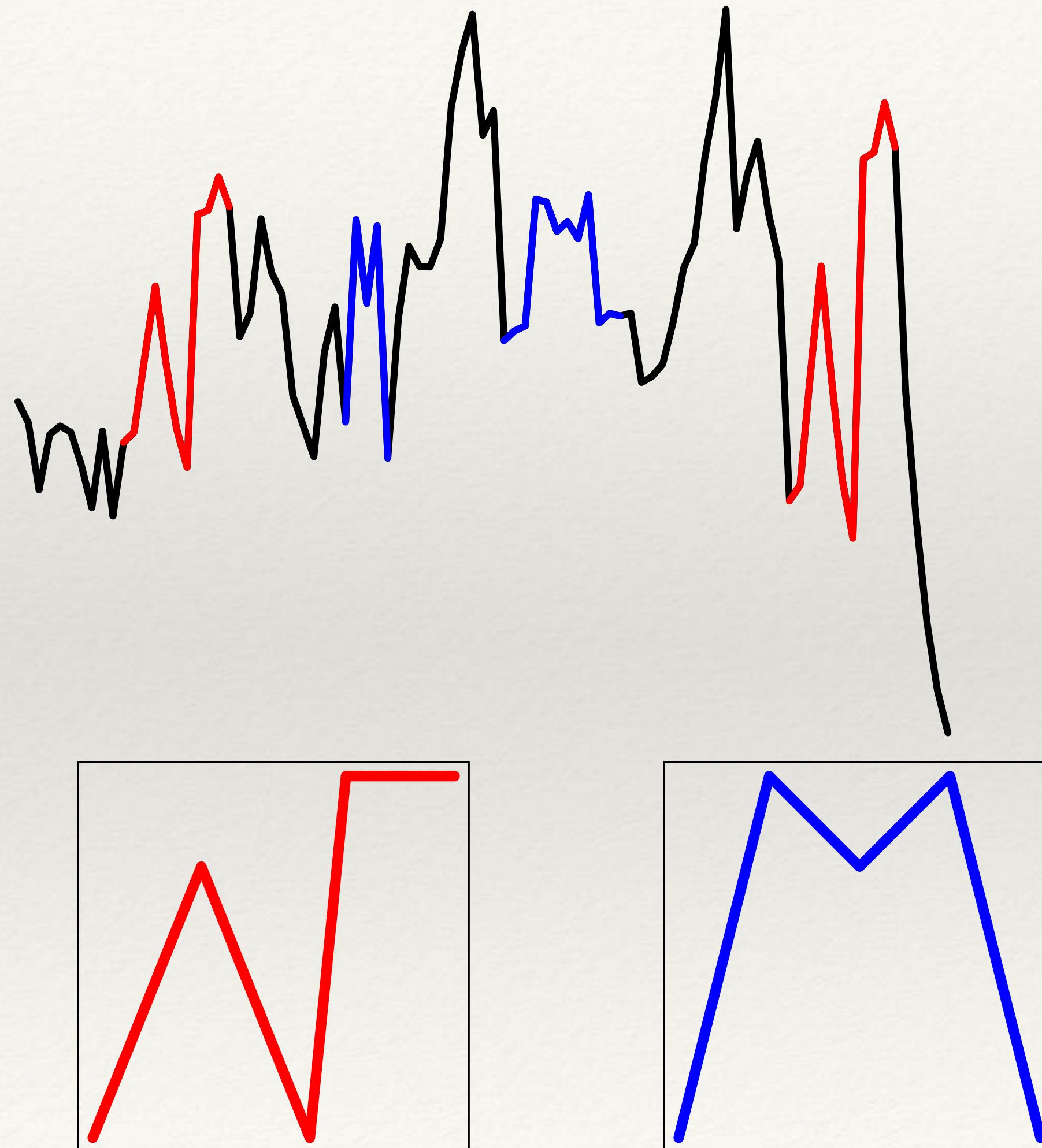
# Forecasting

---

- ❖ **Backward forecasting**
- ❖ **Live forecasting**

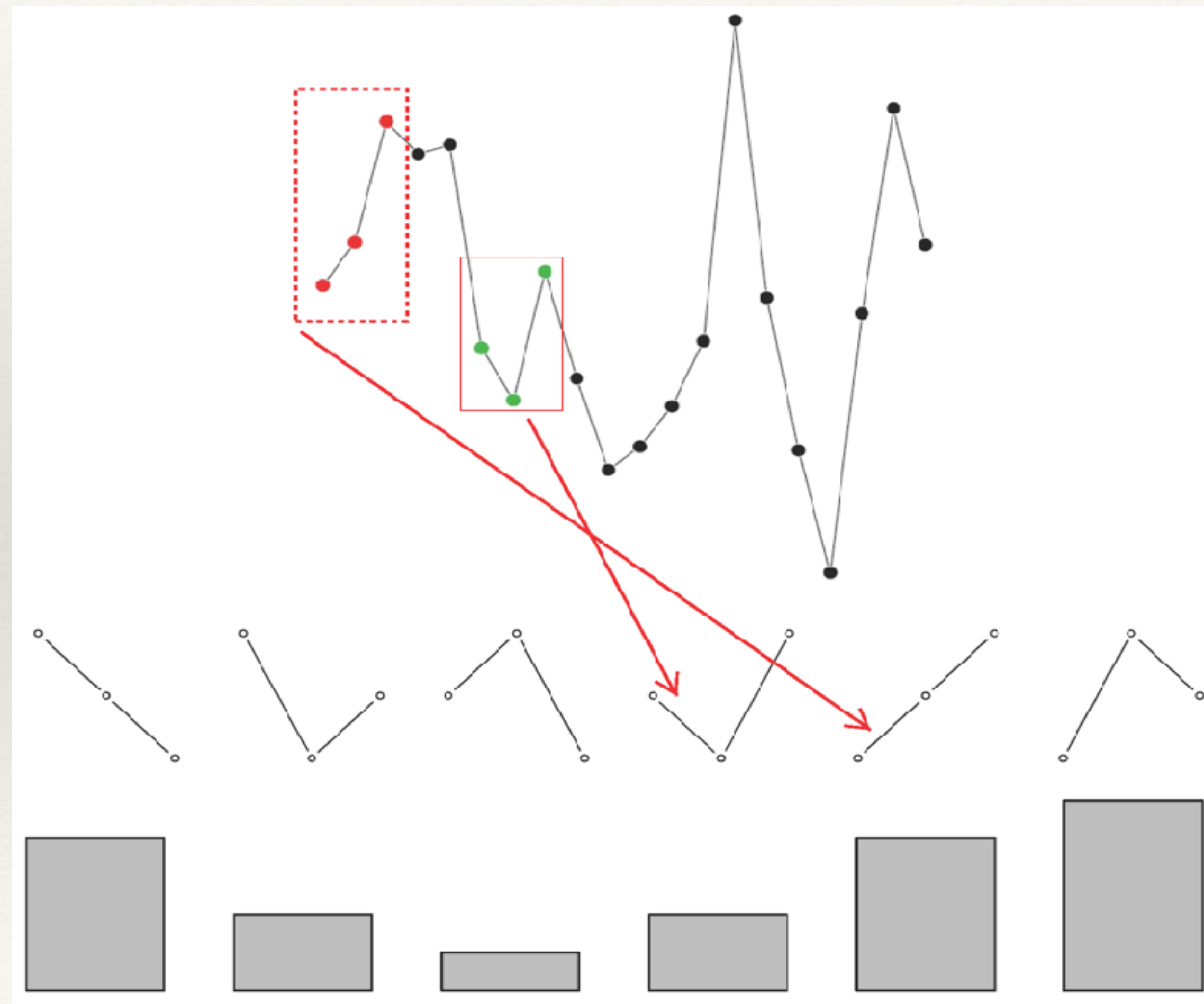
# Theoretical implications 1

## Prototypes for theory building

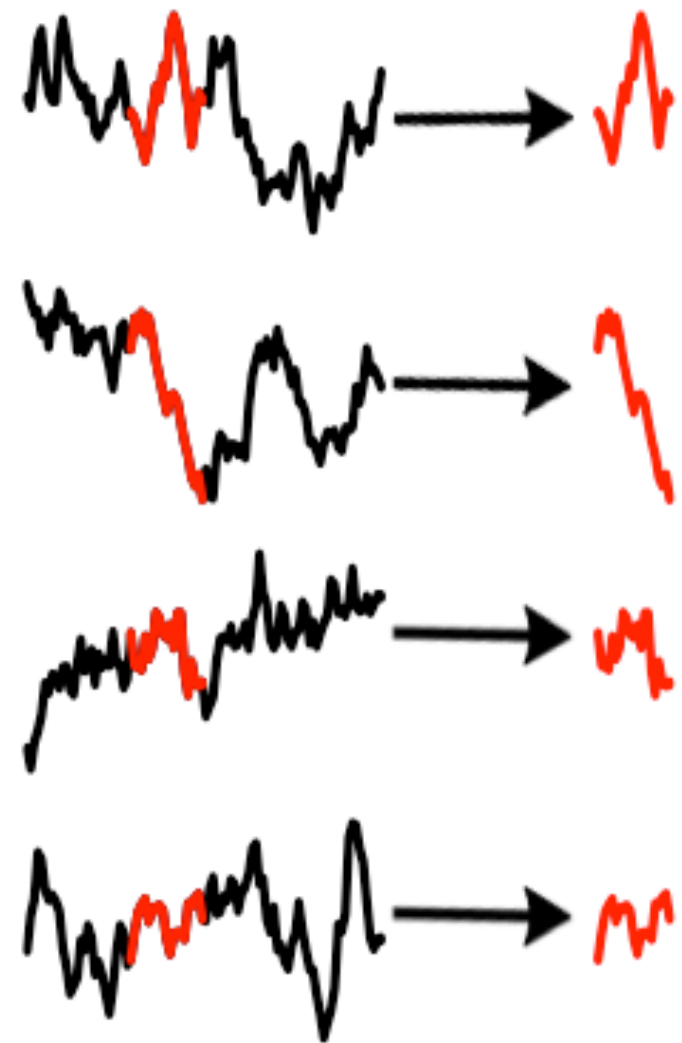


# Theoretical implications 2

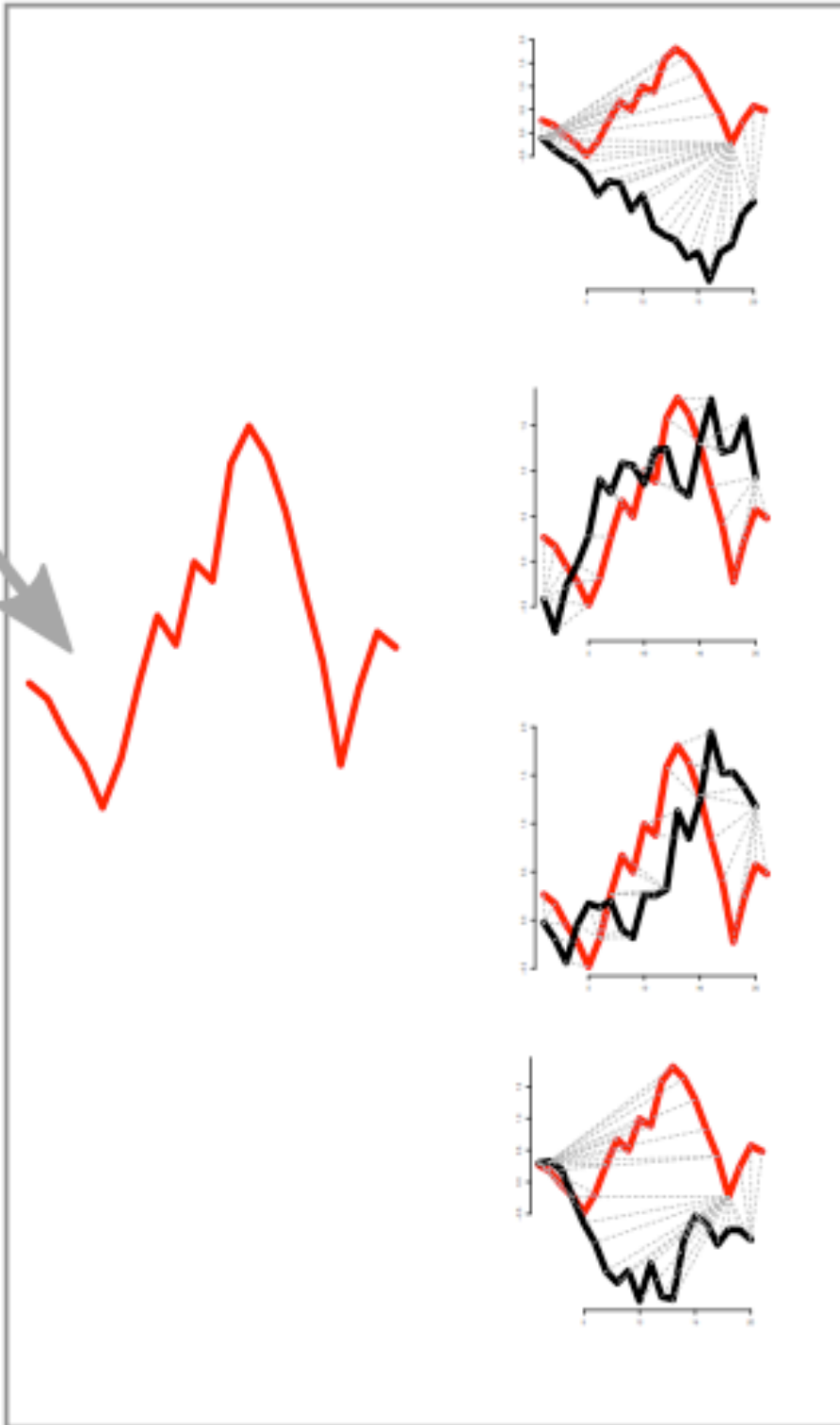
Fundamental limits to the predictability of conflict events?



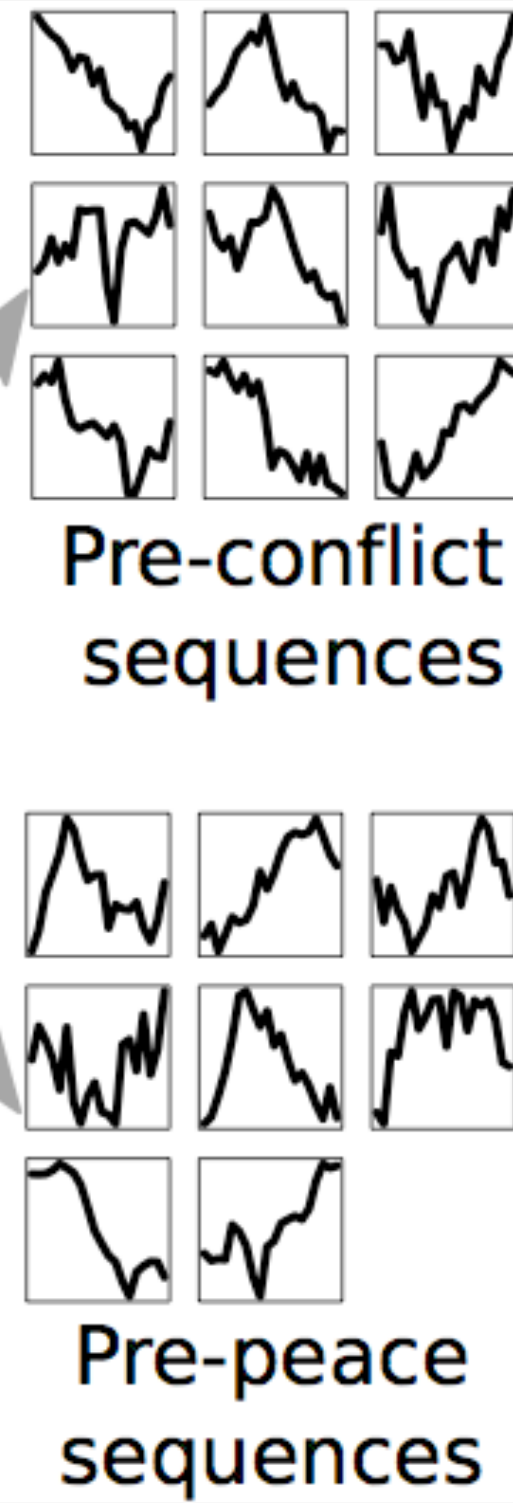
### Data acquisition & processing



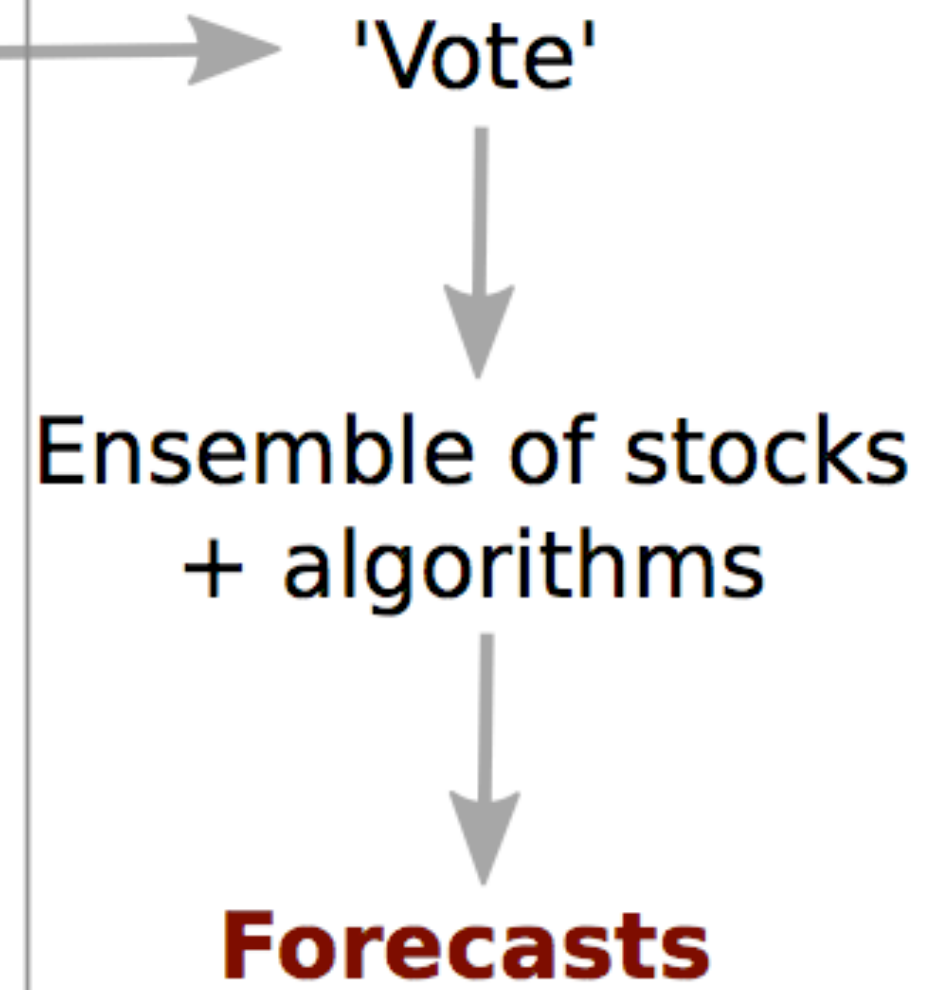
### Distance measures



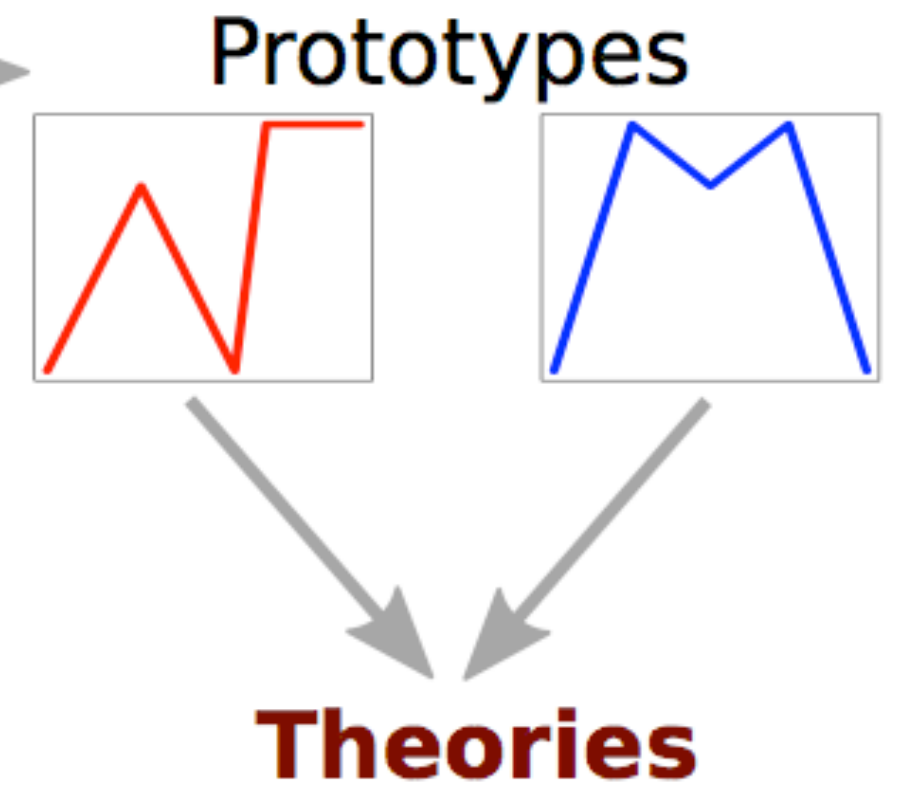
### Classification of series



### Ensemble and forecast



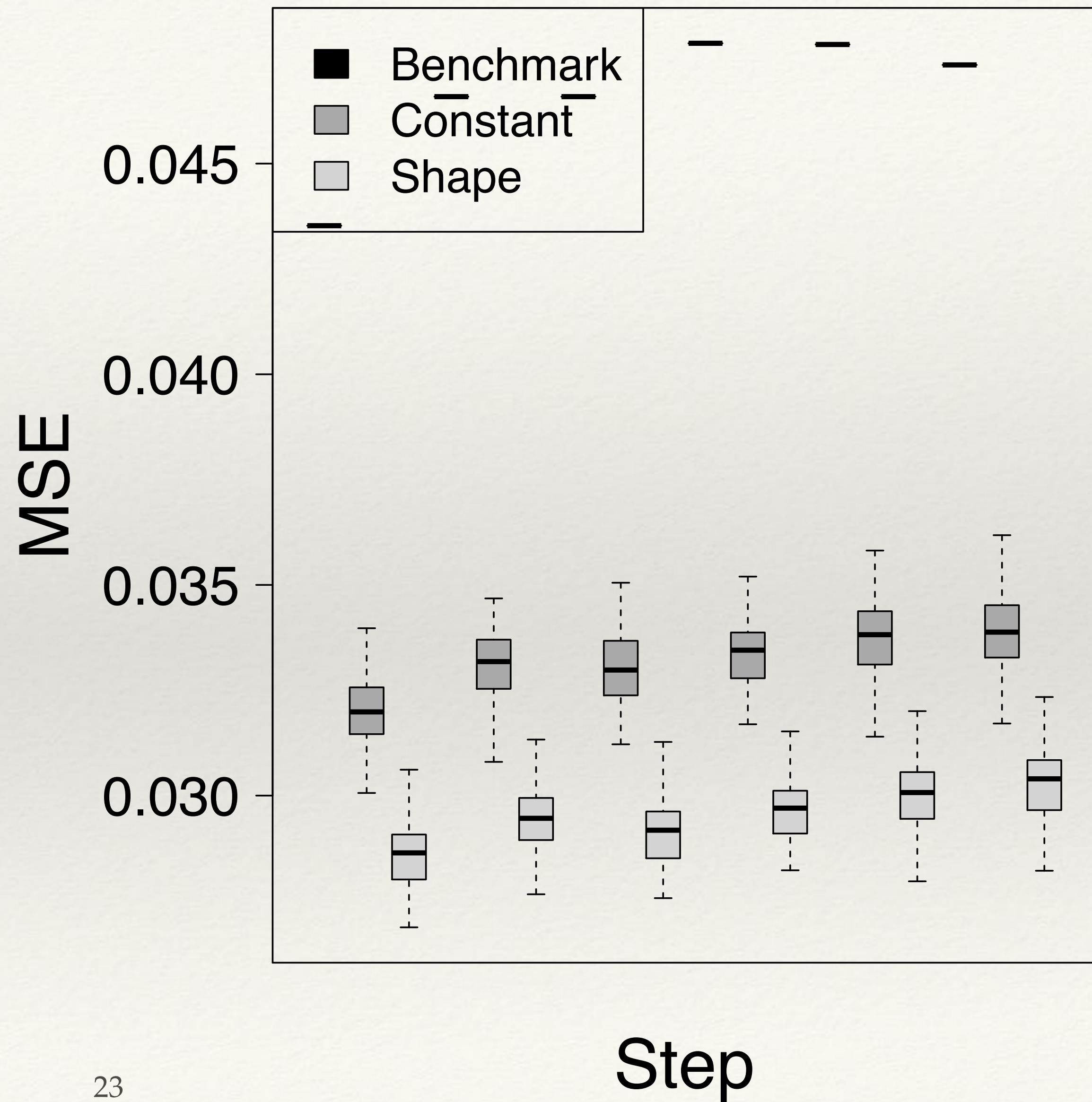
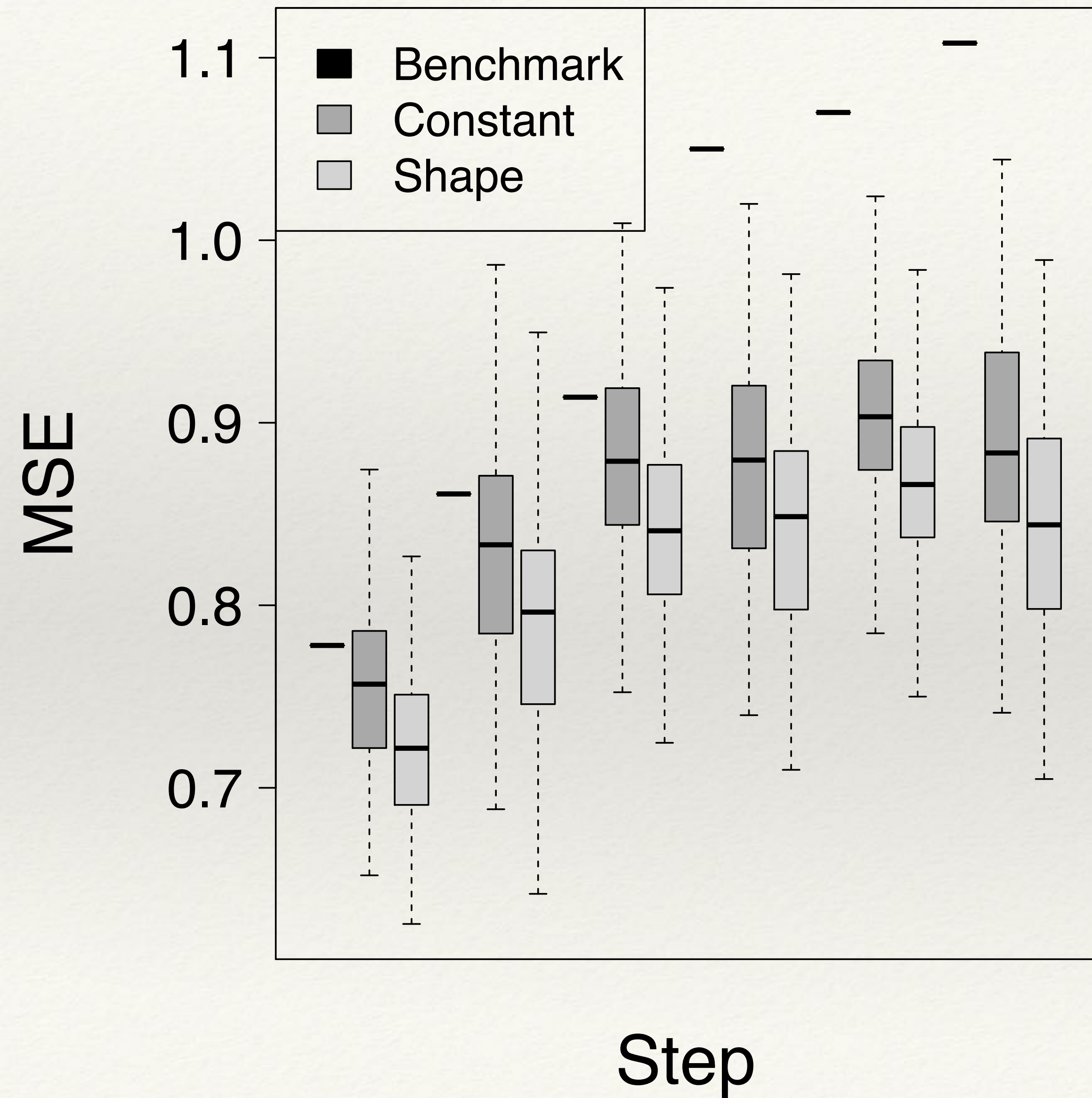
### Theoretical implications



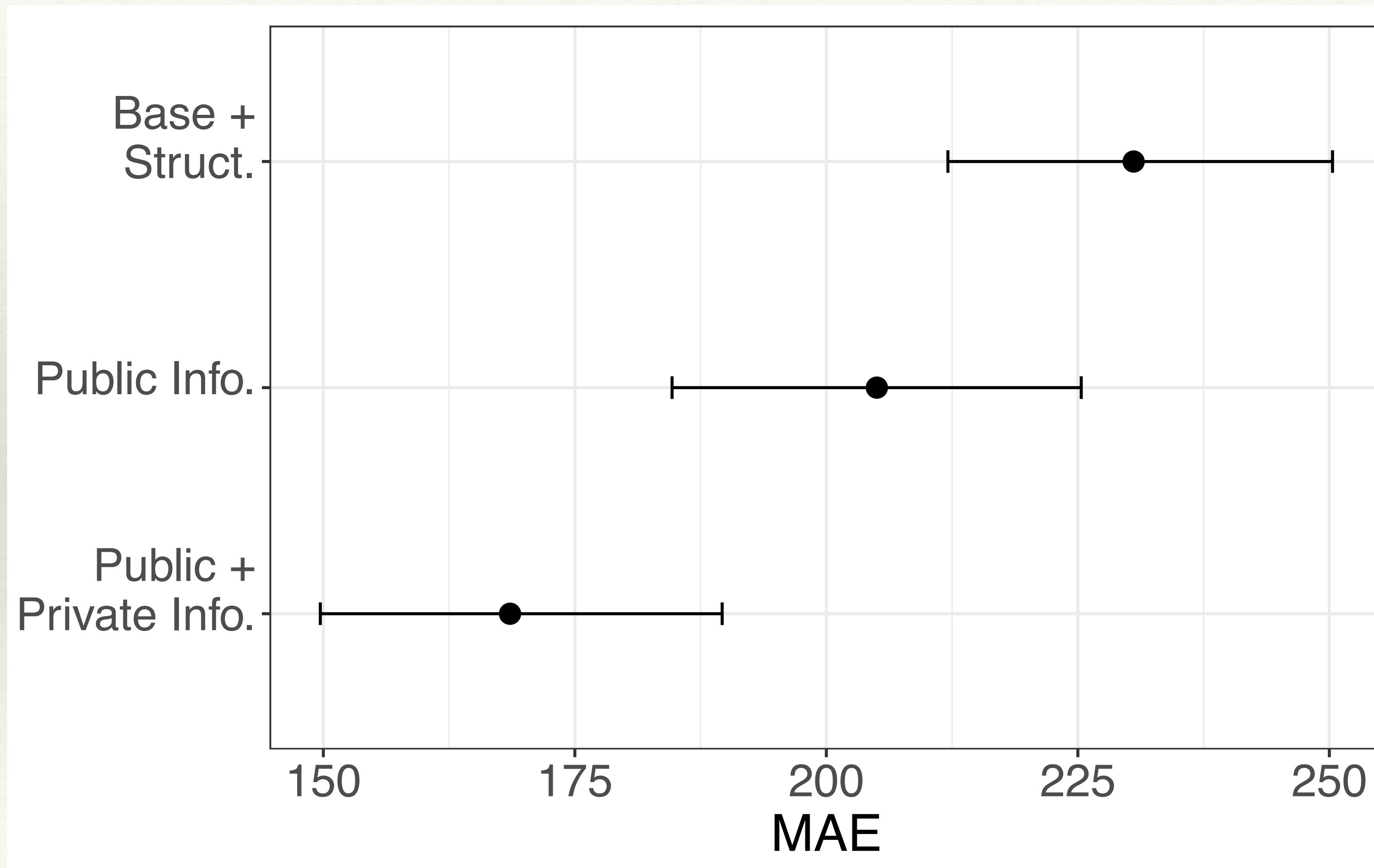


# Some Preliminary results

# Views Competition



# Private and Public Information





---

# Initial projects (ongoing)

---

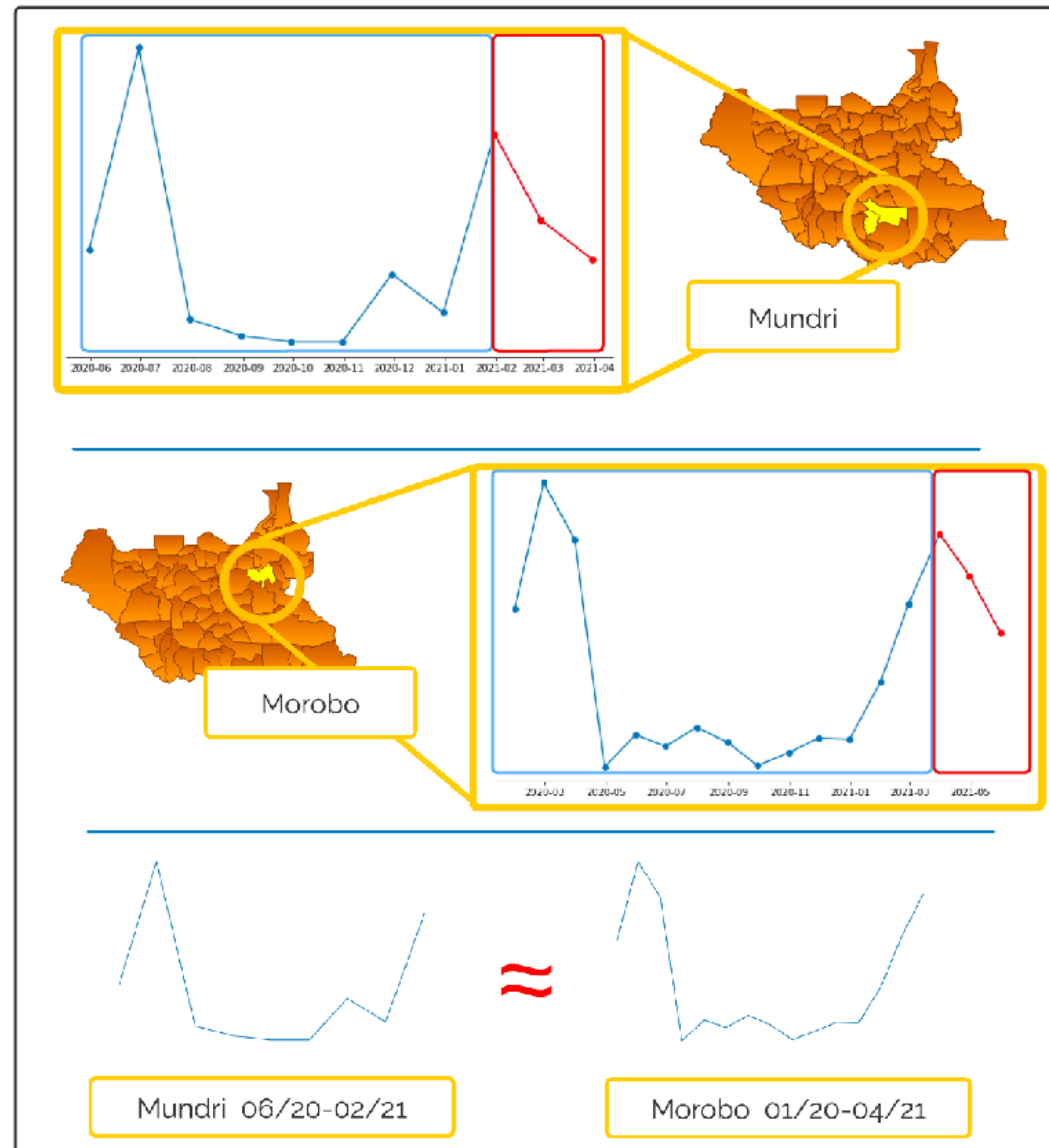
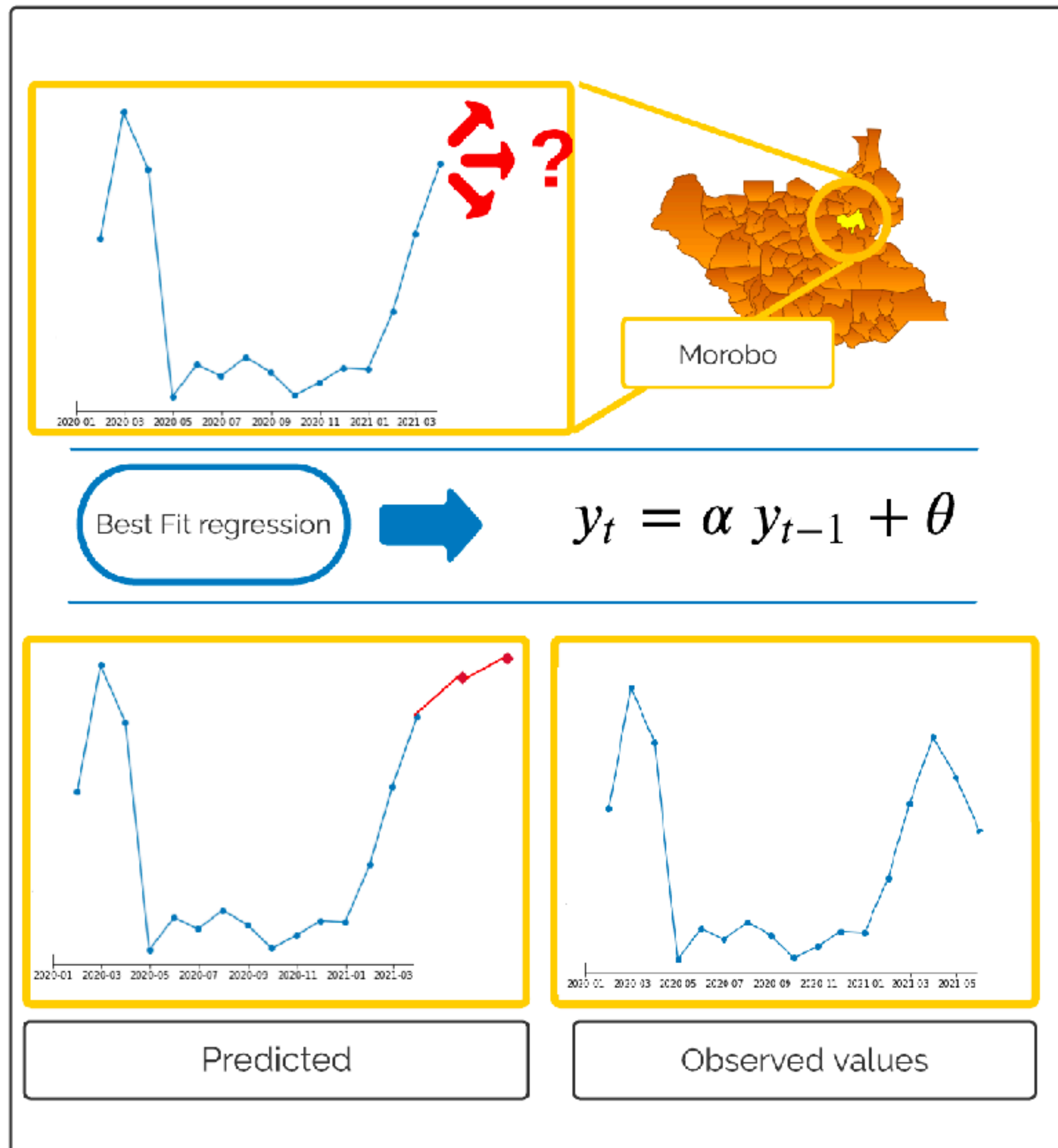
1. Temporal patterns in migration flows
2. Temporal patterns in protest data
3. Augmentation for conflict time series
4. Reducing uncertainty in conflict events using satellite data
5. Dynamic Synthetic Controls (if time allows)

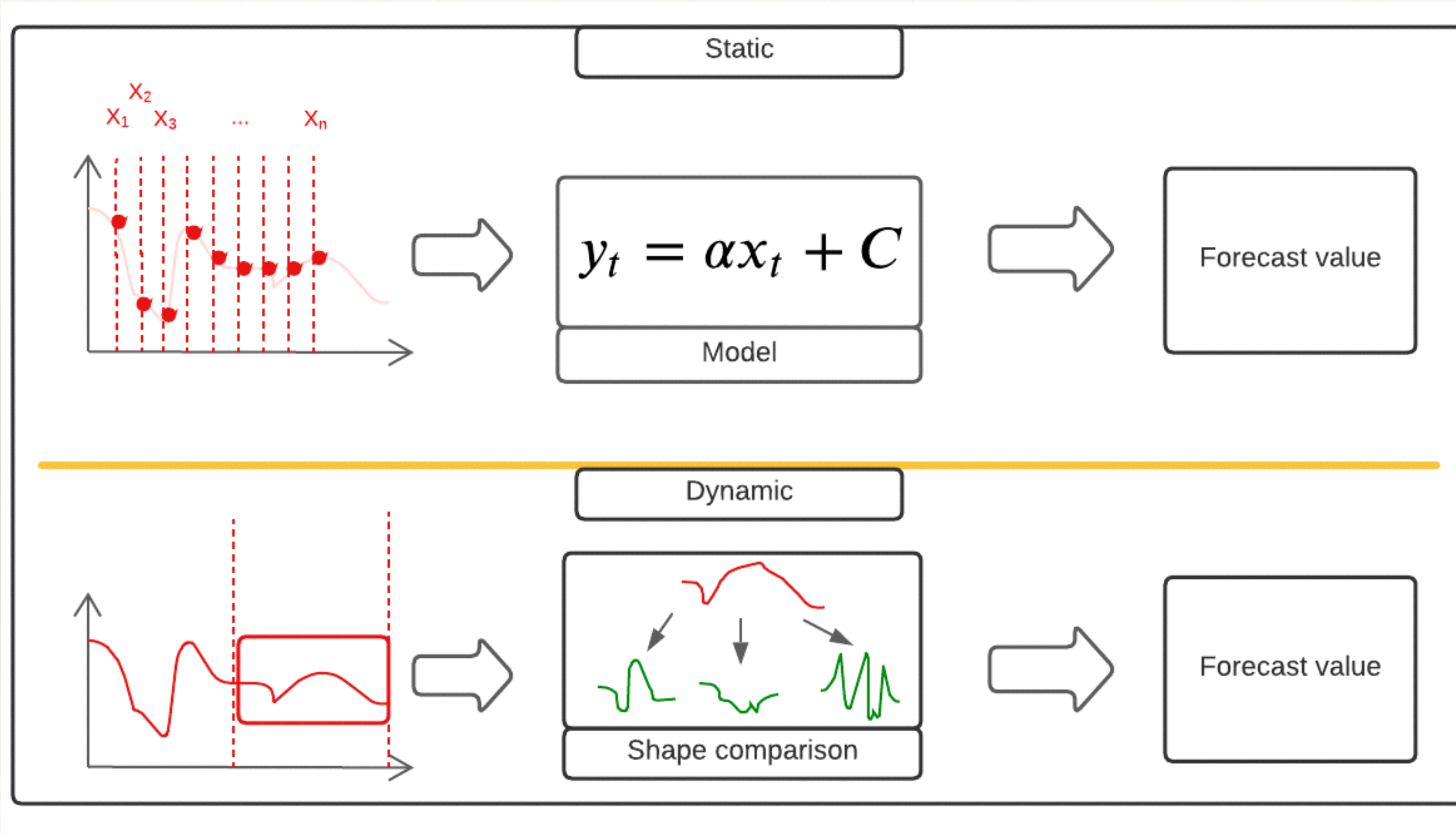
# Project 1

## Temporal patterns in migration flows

Thomas Schincariol & Thomas Chadeaux

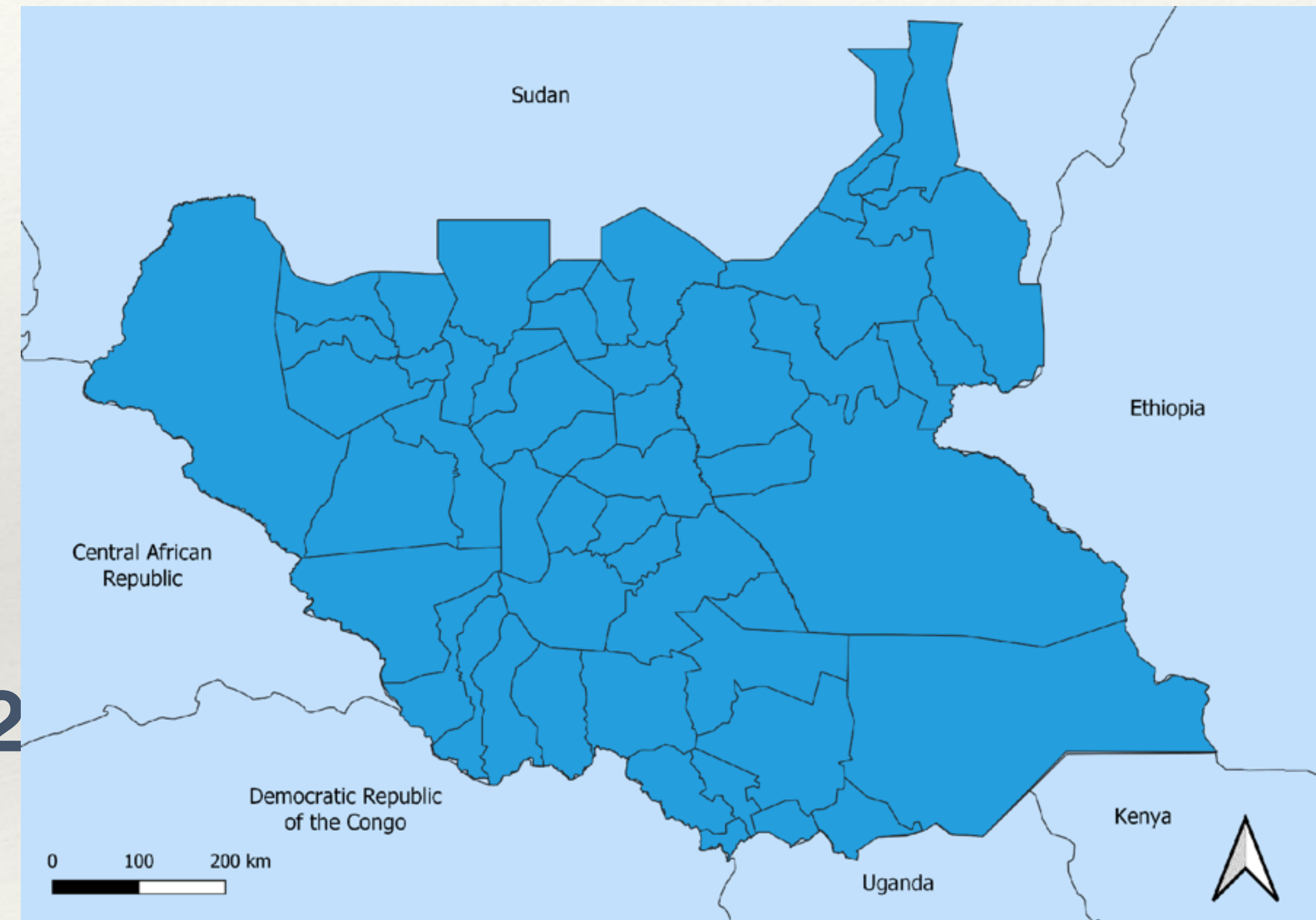
# Example : Migration flow in Morobo - South Sudan





# Study zone - South Sudan

- ❖ **48 regions**
  - ❖ Adm - 2 level
  - ❖ Regional merging if value < 10  
→ **72 to 48**
- ❖ **January 2020 to September 2021**
- ❖ **21 point time series**



---

# Data Source

---

- ❖ **IOM-DTM flow monitoring surveys**
  - ❖ **31 key transit points within South Sudan and at its borders**
  
- ❖ **Focus on population leaving their region**
  - ❖ **Push** in the “push-pull” model
  - ❖ **Mostly economic, conflict or climate motivations**

# Method : Benchmark of autoregressive models

## Basic Model

Autoregressive model

$$y_t = c + \alpha * y_{t-1}$$

No parameters

→ no optimization

VS

## Common Literature model

ARIMA model

$$y'_t = c + \sum_{i=1}^p \alpha_i * y'_{t-i} + \sum_{j=1}^q \beta_j * \epsilon_{t-j}$$

3 parameters (p,d,q)

VS

## Dynamic model

**ARIMAX**

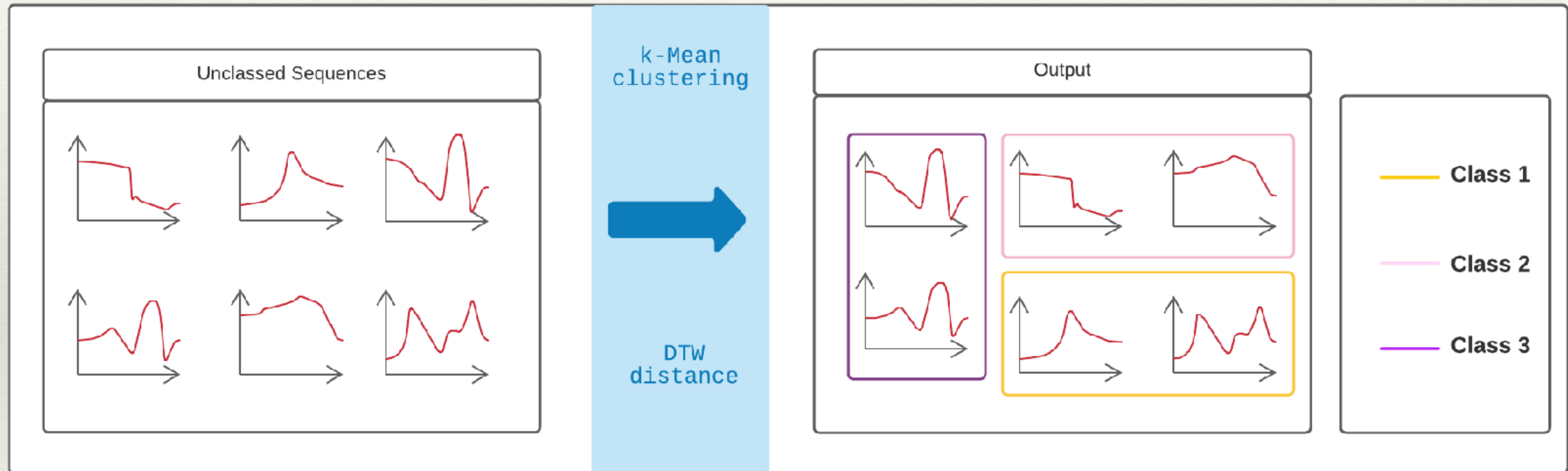
$$y'_t - c + \sum_{i=1}^p \alpha_i * y'_{t-i} + \sum_{j=1}^q \beta_j * \epsilon_{t-j} + \gamma_c$$

$$\gamma_c = \begin{cases} \gamma_1, & \text{class}_1 = 1 \\ \gamma_2, & \text{class}_2 = 1 \\ \gamma_3, & \text{class}_3 = 1 \end{cases}$$

**5 parameters :**

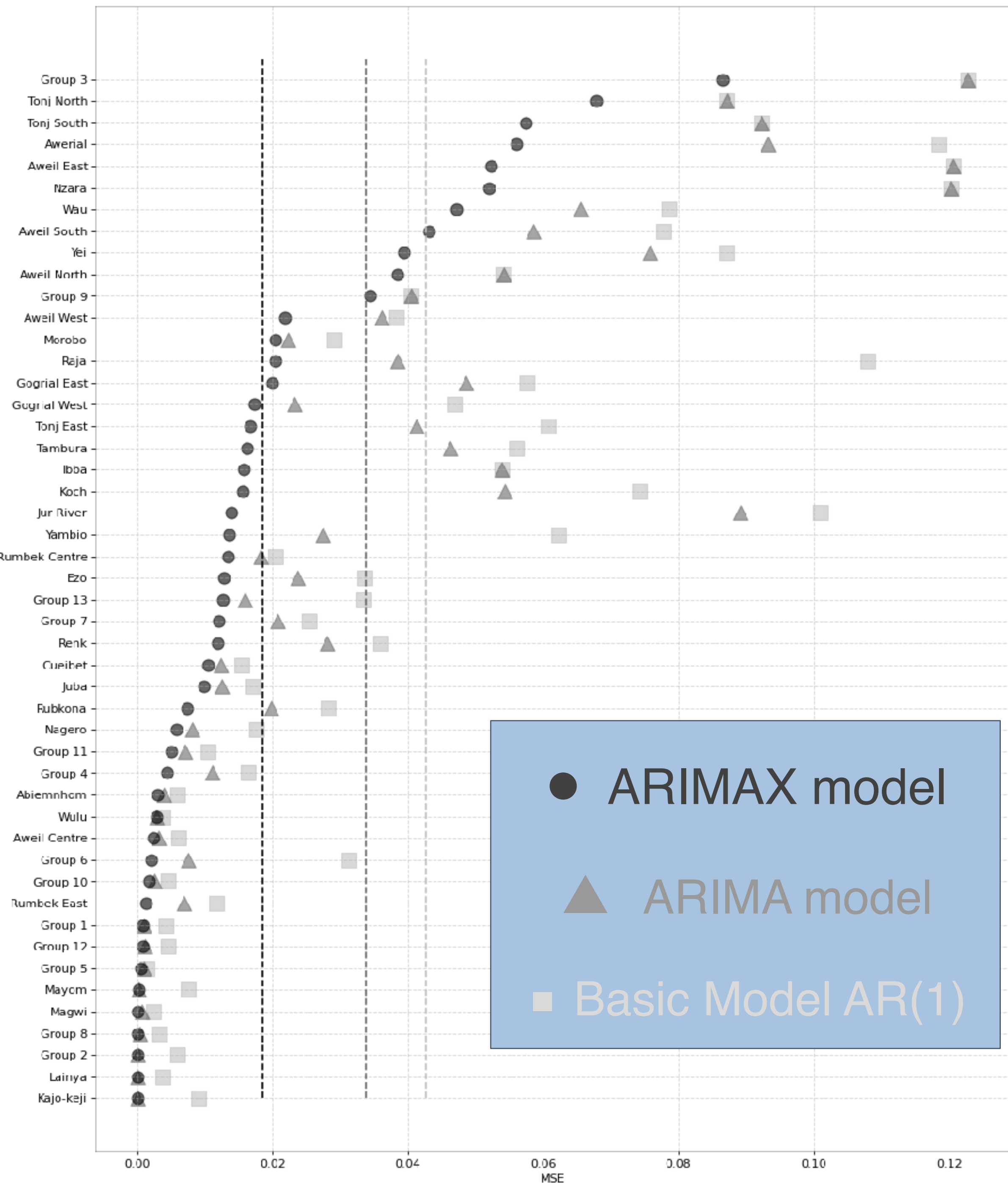
- 3 ARIMA (p,d,q)
- 2 DTW classification

# Method : DTW - Classification application





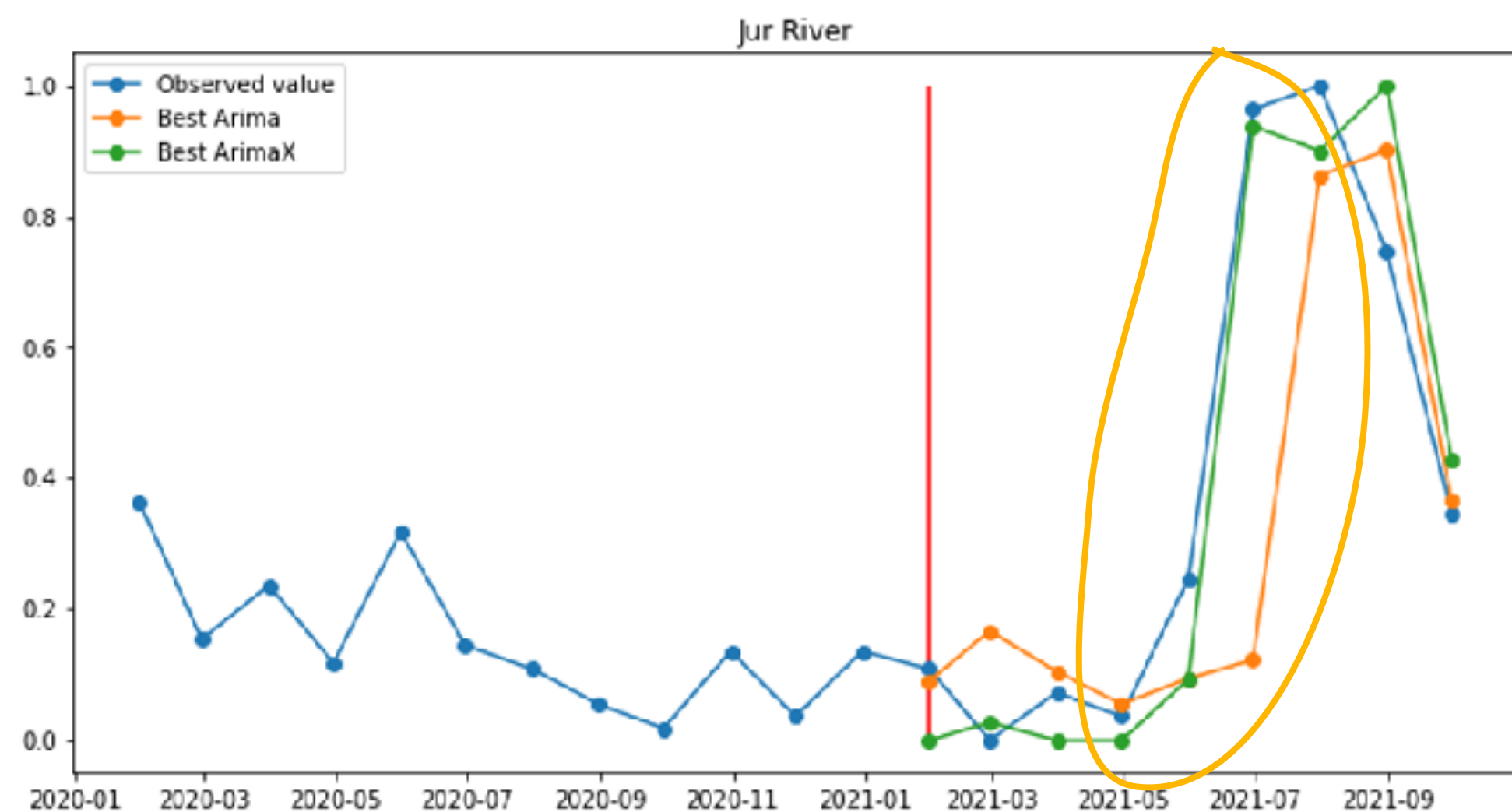
MSE of the models



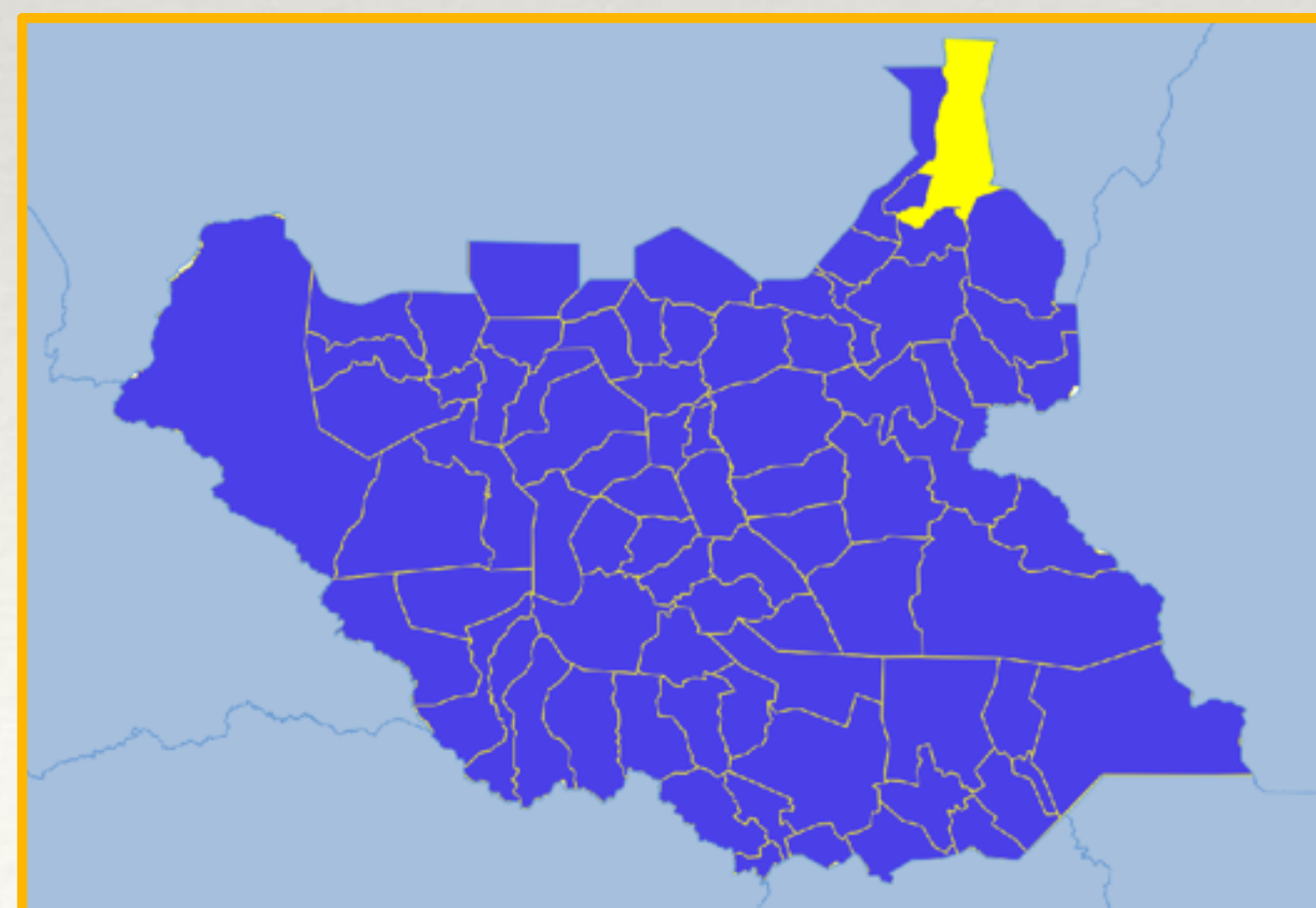
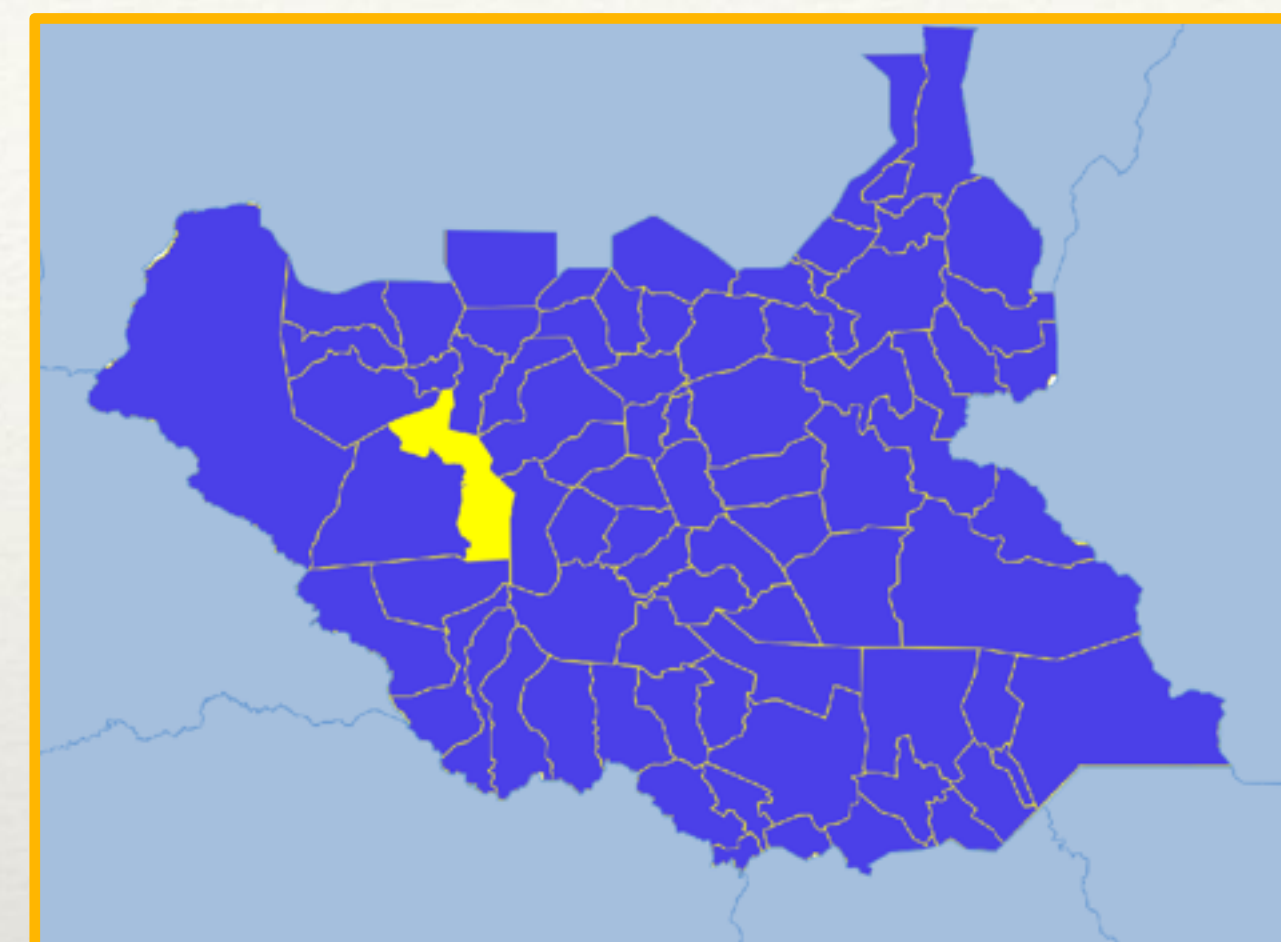
# MSE Plots

Mean value of MSE  
for each South  
Sudan regions

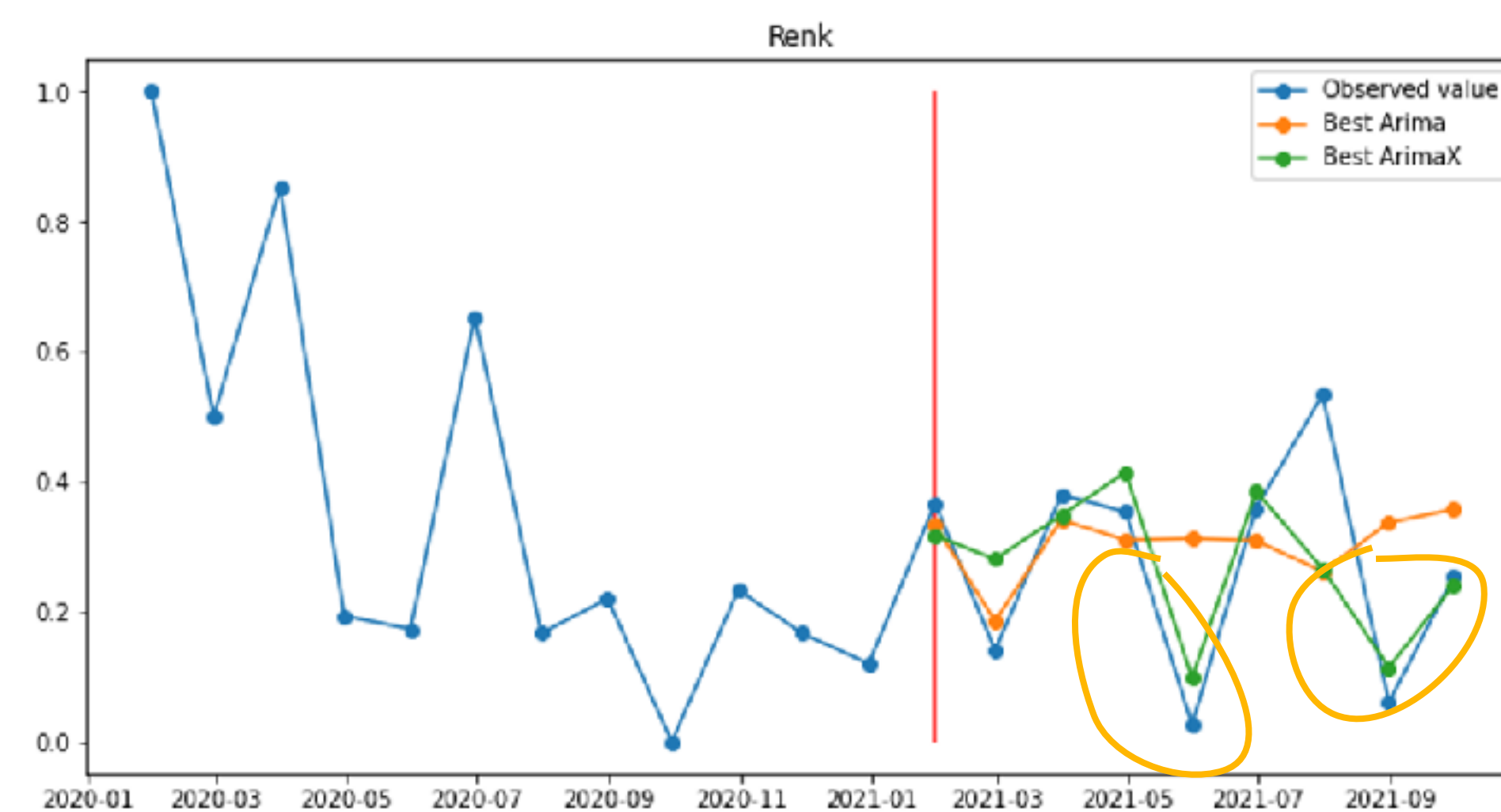
# Region forecast examples : Jur River and Renk



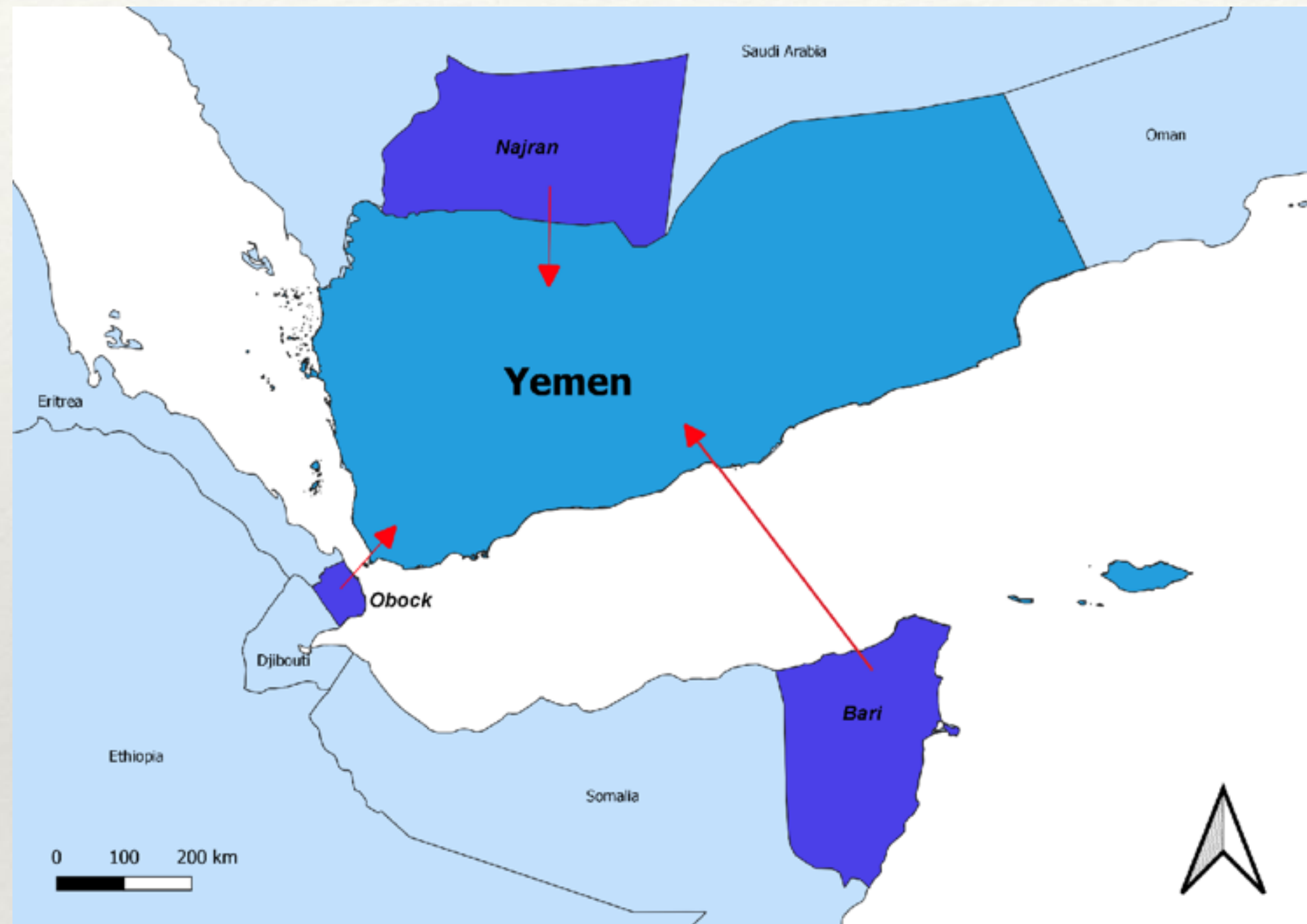
Jur River



Renk



# Beyond South Sudan : Yemen and Pakistan

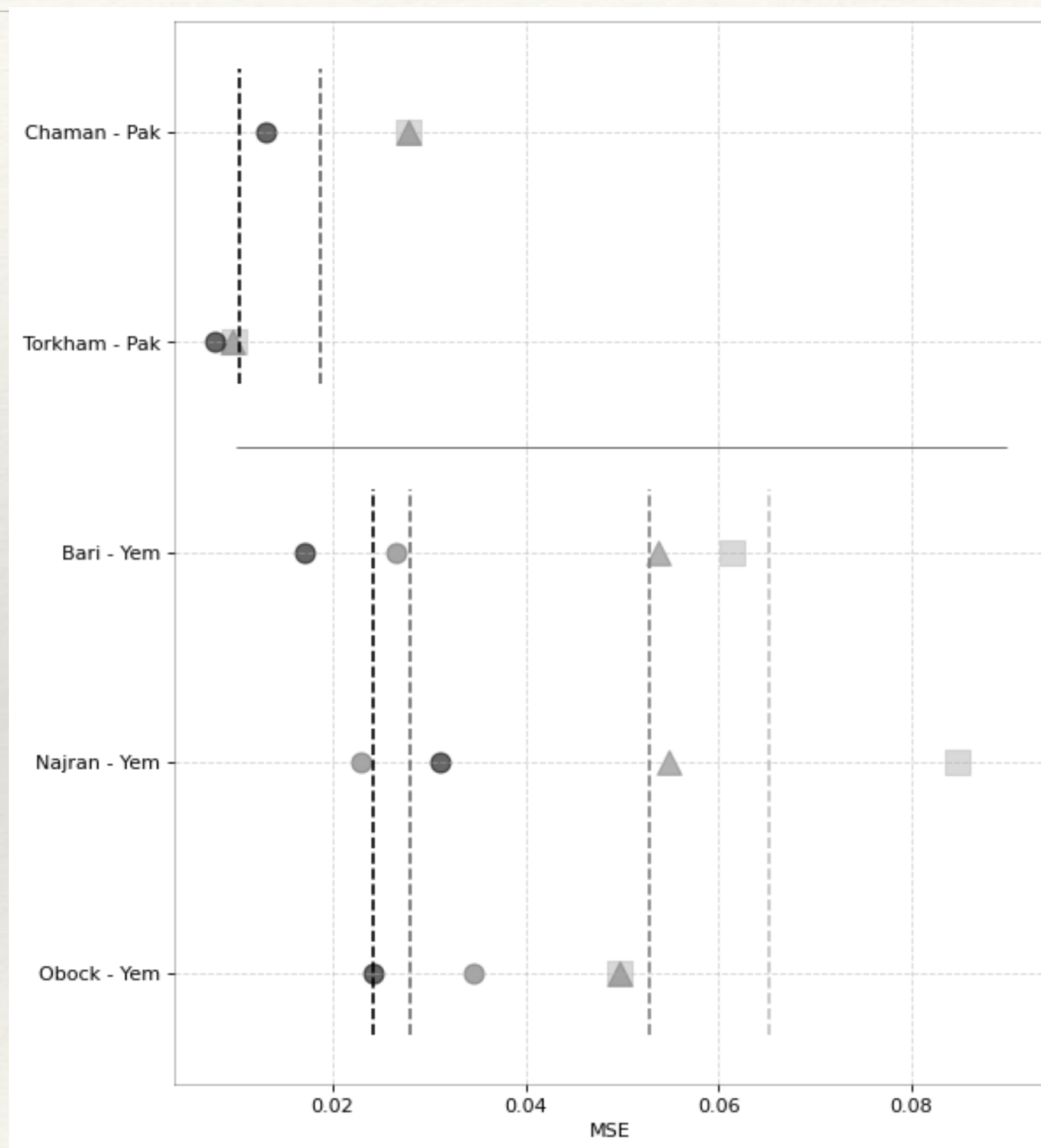


- ❖ *Bari (Somalia), Obock (Djibouti), Najran (South Arabia)*
- ❖ *49 months time series (Jan 2018-Jan 2022)*



- ❖ *Torkham and Chaman points*
- ❖ *52 weeks time series*

# Beyond South Sudan : Yemen and Pakistan



## Mean value of MSE for Pakistan and Yemen

- ARIMAX model
- ARIMAX model (*South Sudan*)
- ▲ ARIMA model
- Basic Model AR(1)

---

# Conclusion

---

- ❖ **Good results in South Sudan, Yemen and Pakistan**
- ❖ **Apply the method to other topics**
  - ❖ **Climate**
  - ❖ **Protest**
  - ❖ **Other ideas ?**

# Project 2

## Temporal Patterns in Protest Data

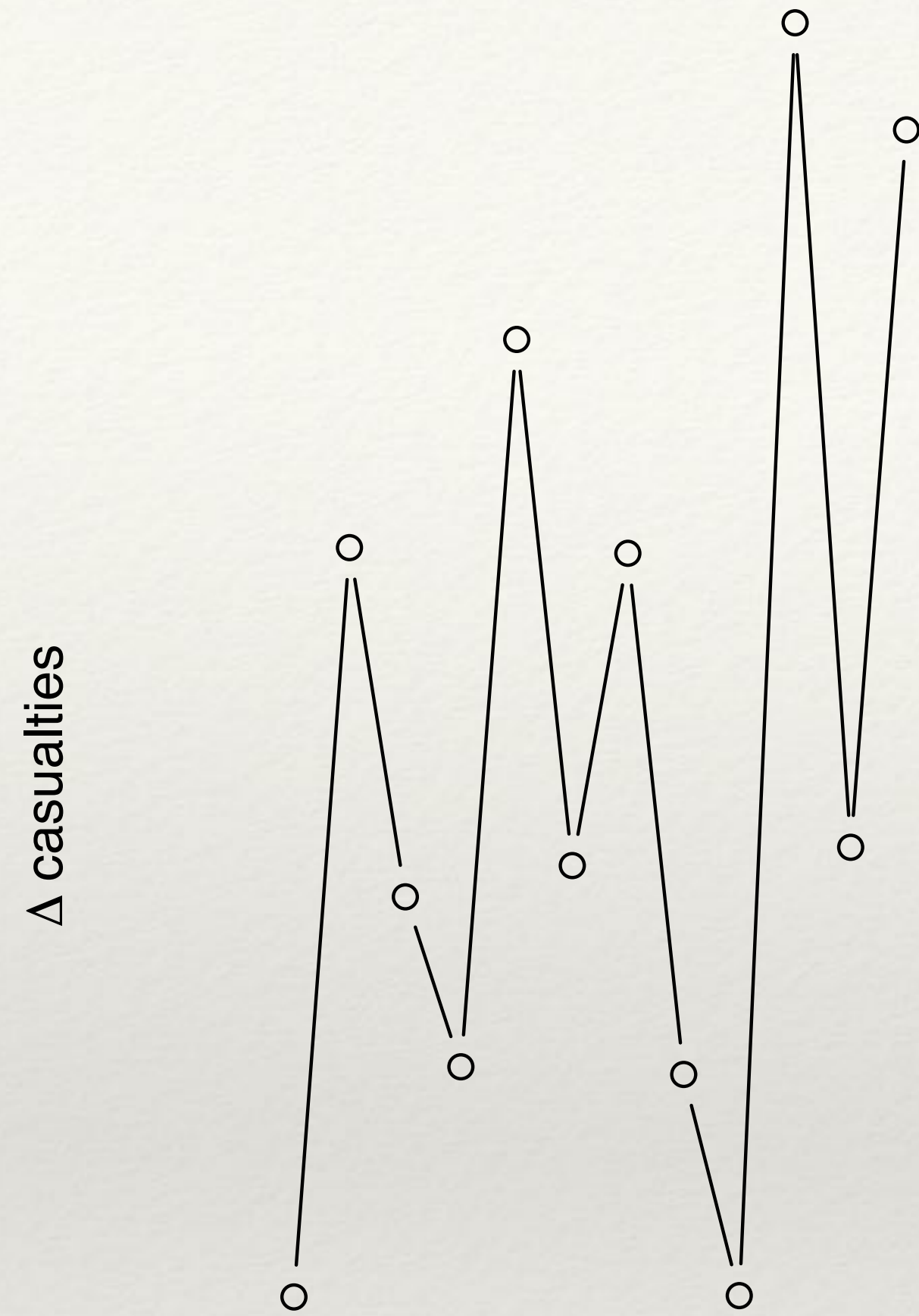
Hannah Frank & Thomas Chadeaux

# Project 3

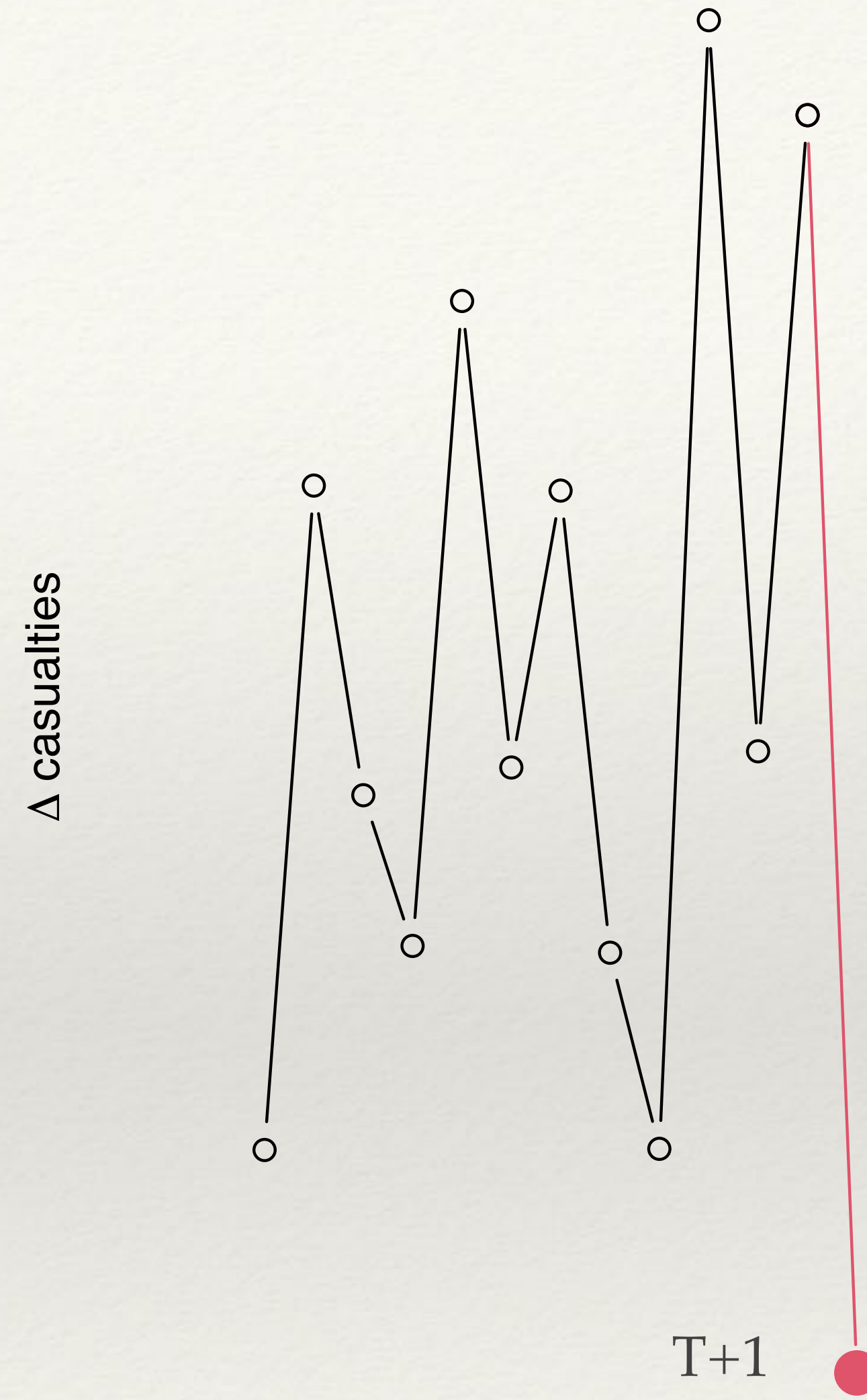
## Augmentation for time series

Thomas Chadeaux

# What we observe



# what we want to predict



Time

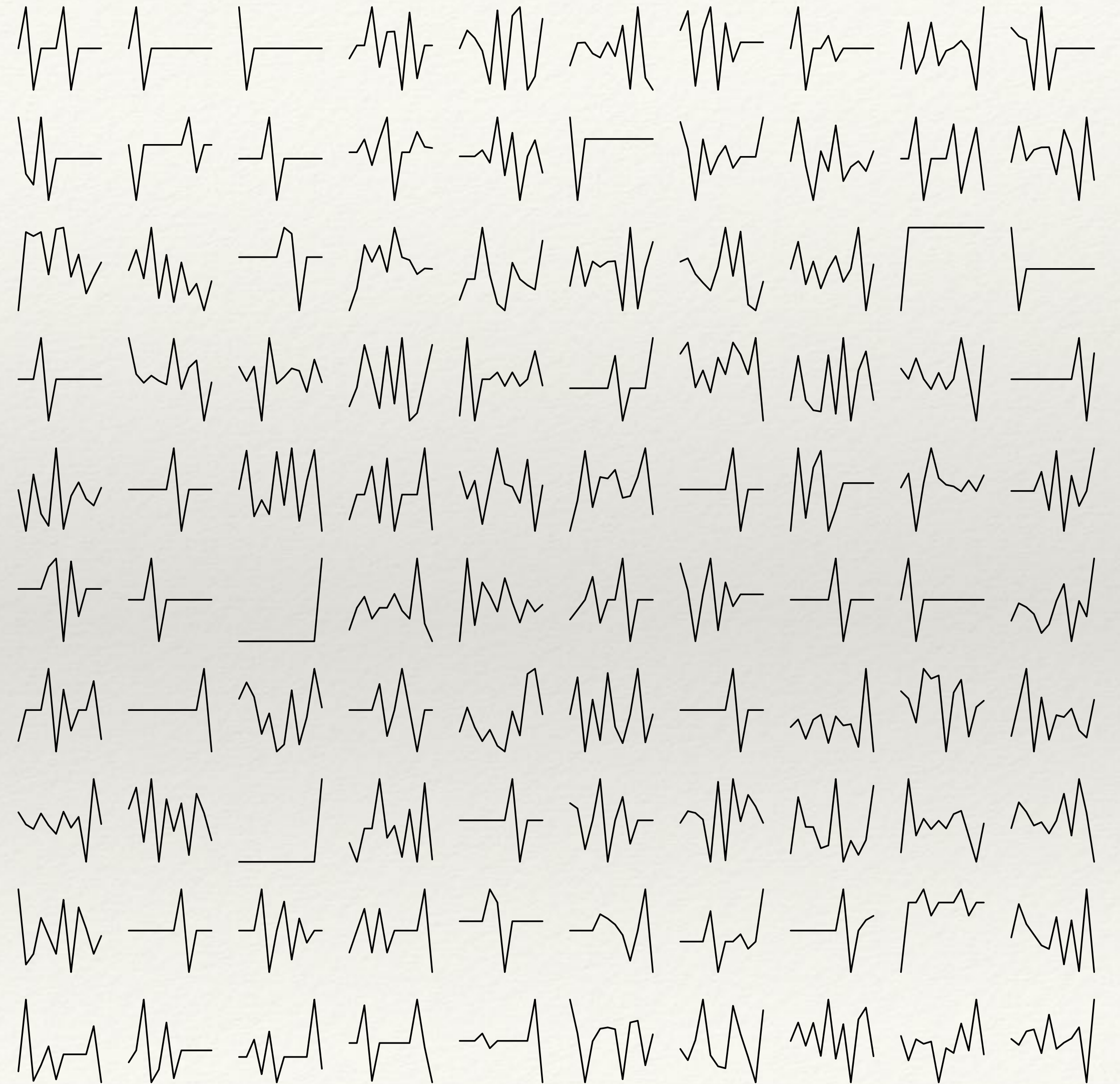
Time



# Learning Set



# Test set



- 
- 
- ❖ **Data is scarce**
    - ❖ **One world**
  - ❖ **Acquiring data is expensive**

---

# Common methods of dealing with limited data in social sciences

---

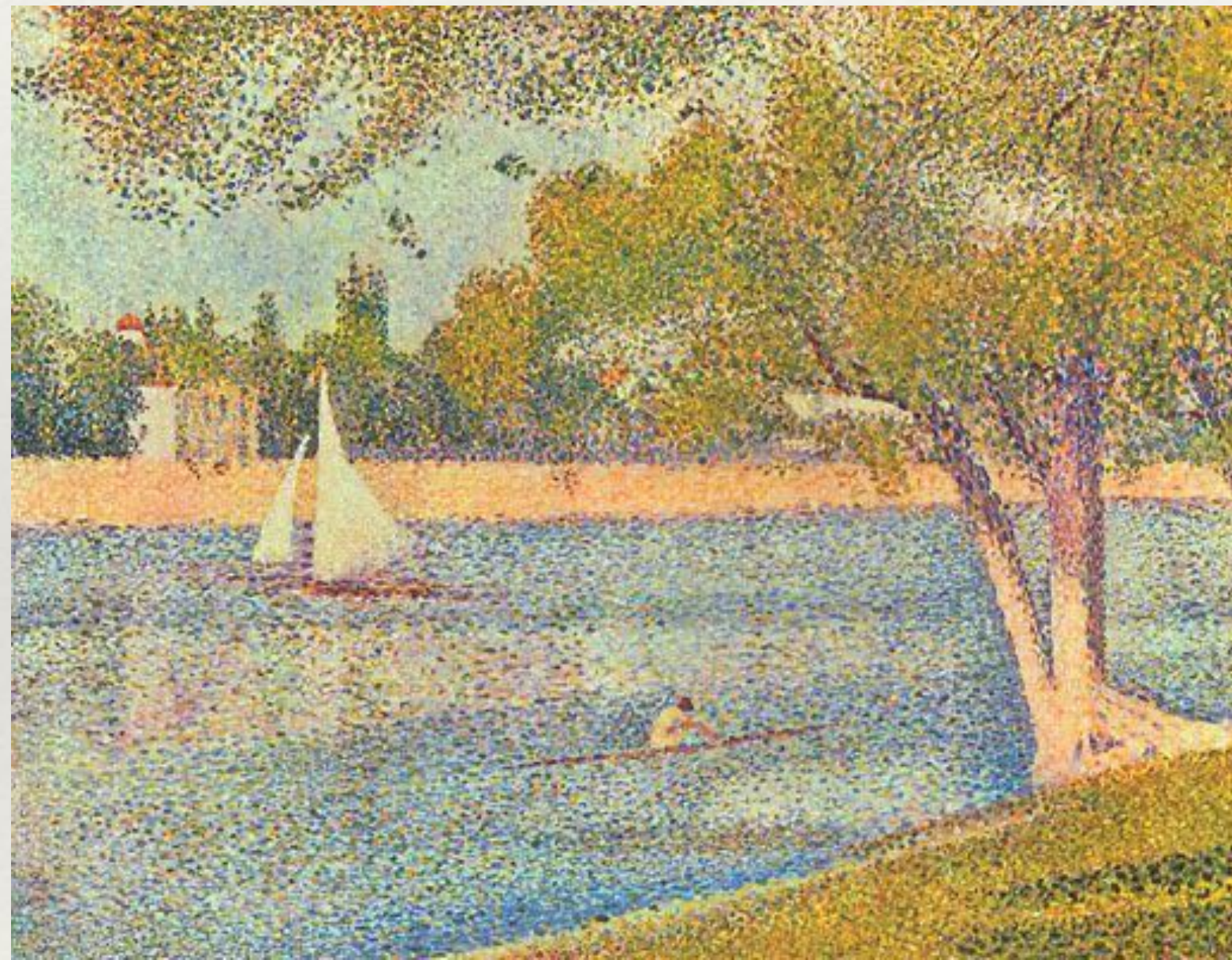
- ❖ **Bootstrapping and Bagging: use subsets of the data.**
- ❖ **Imputation: fill gaps**
- ❖ **Create entirely new dataset: augmentation**

---

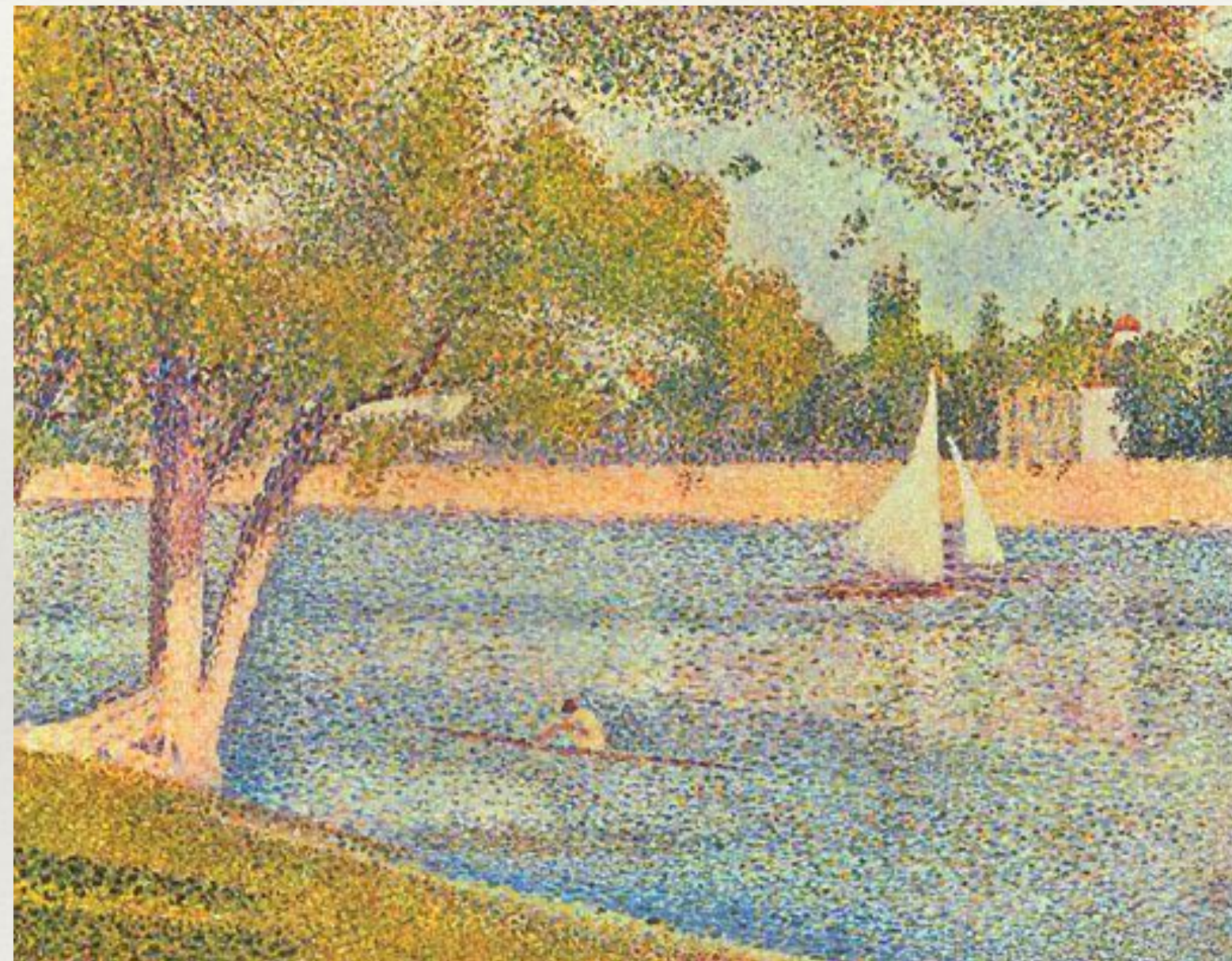
# Augmentation

---

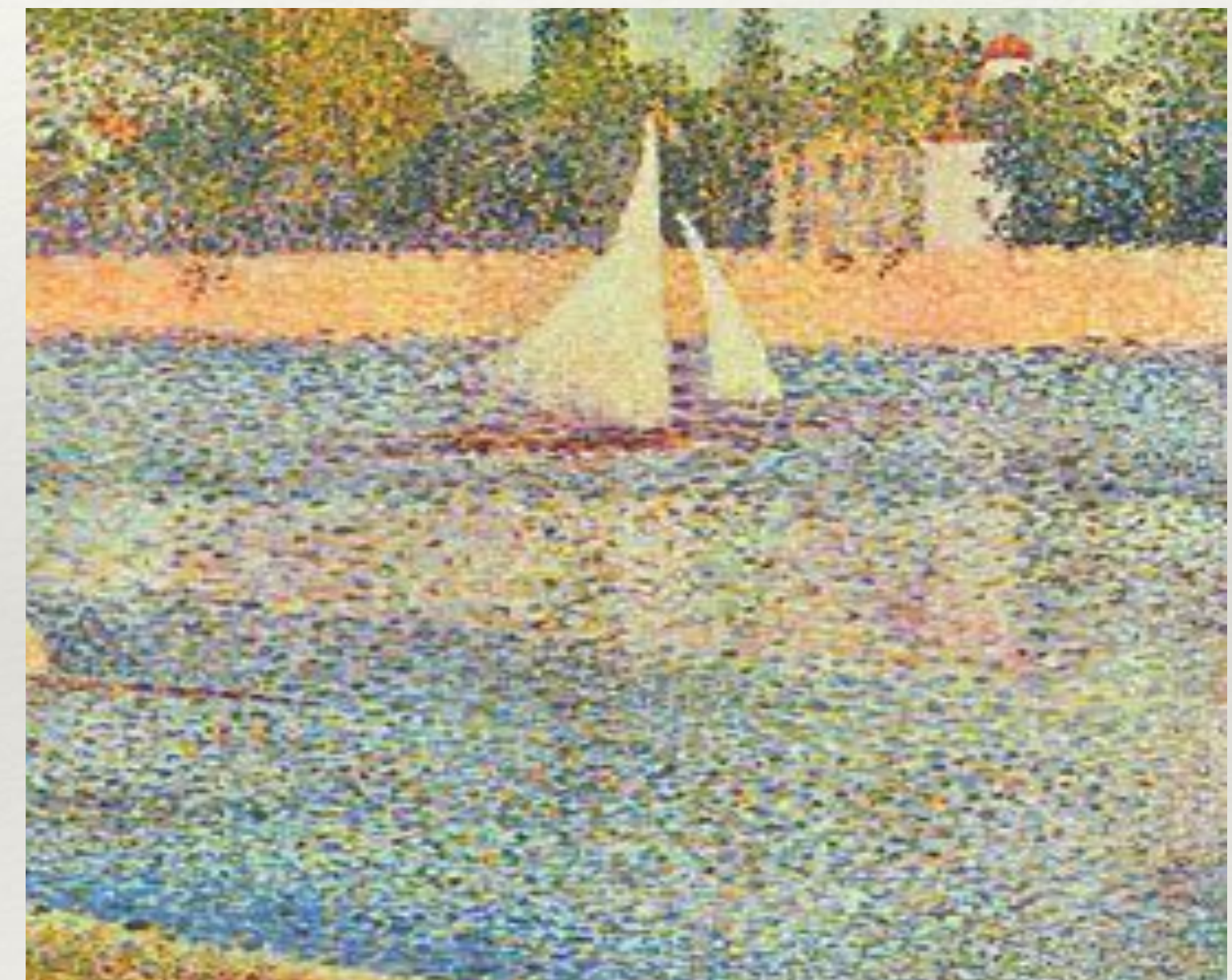
Original



Flipped

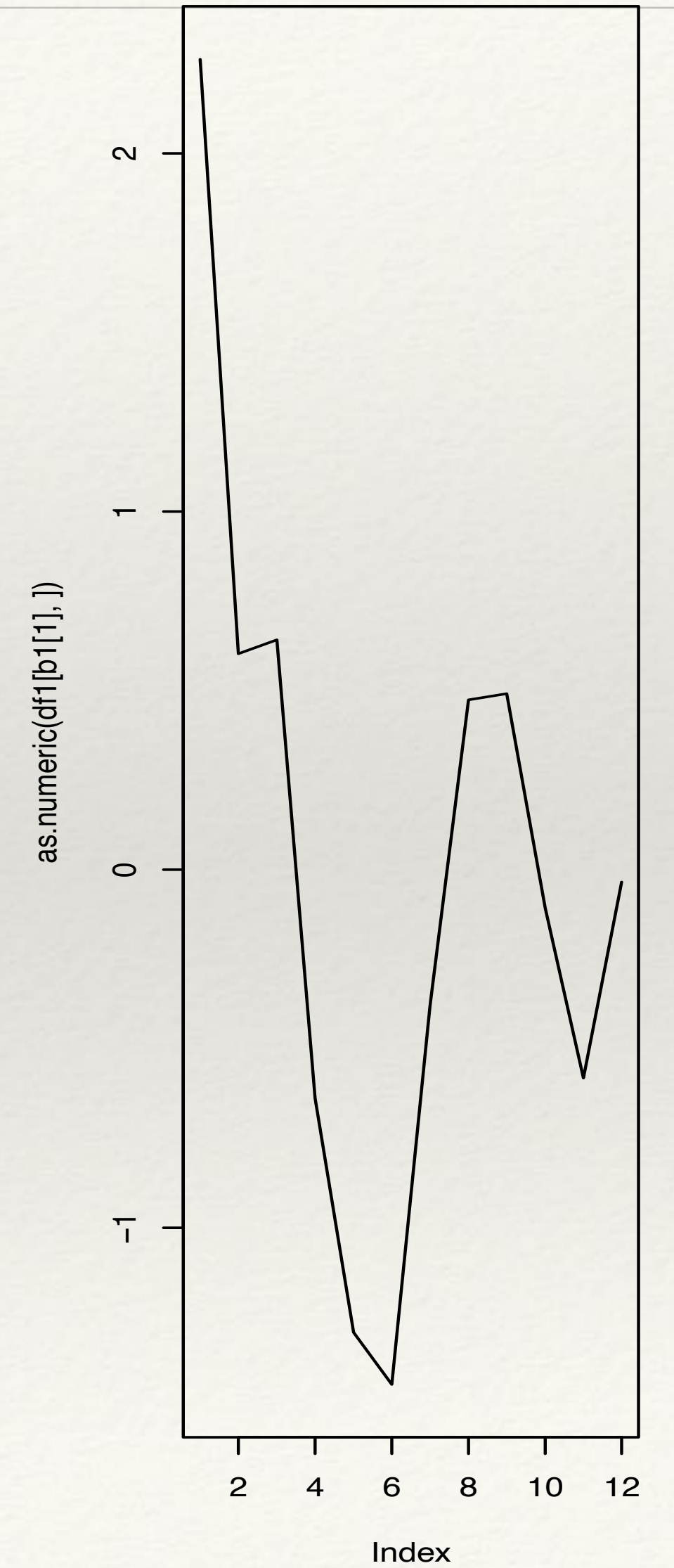
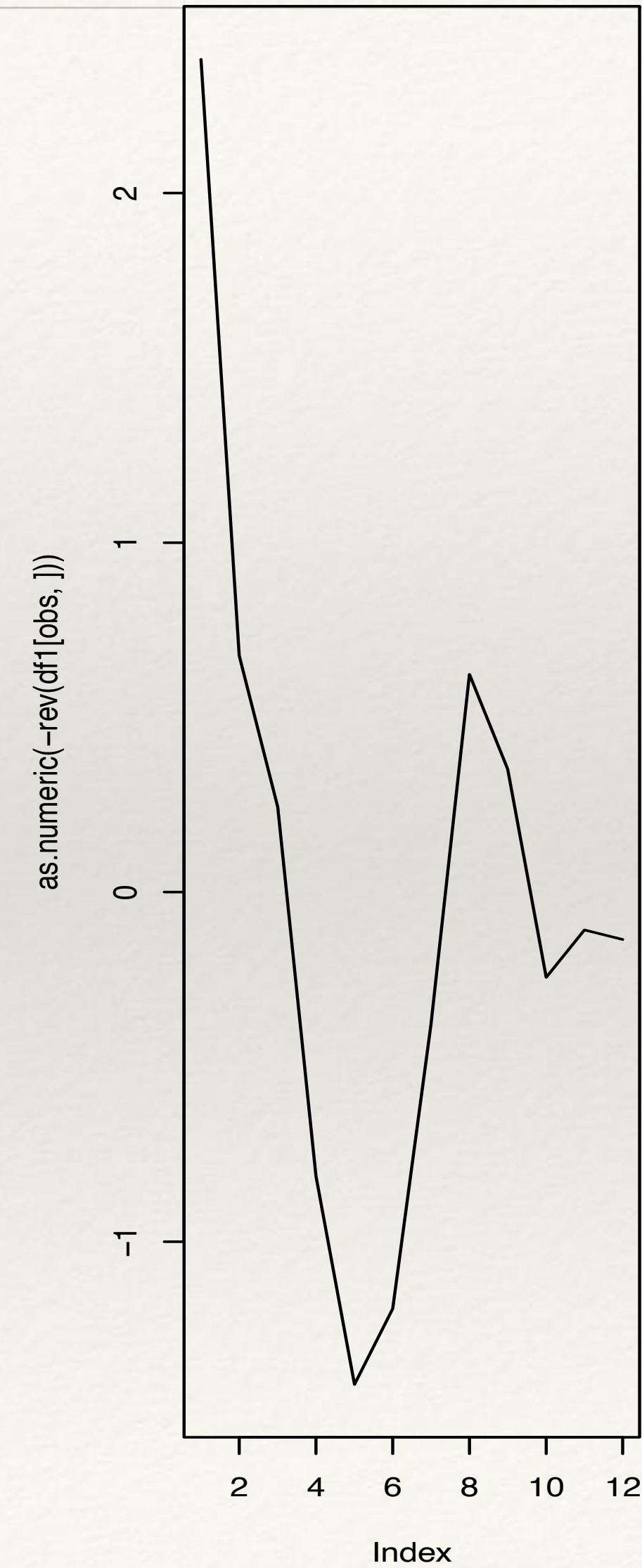


Cropped / stretched

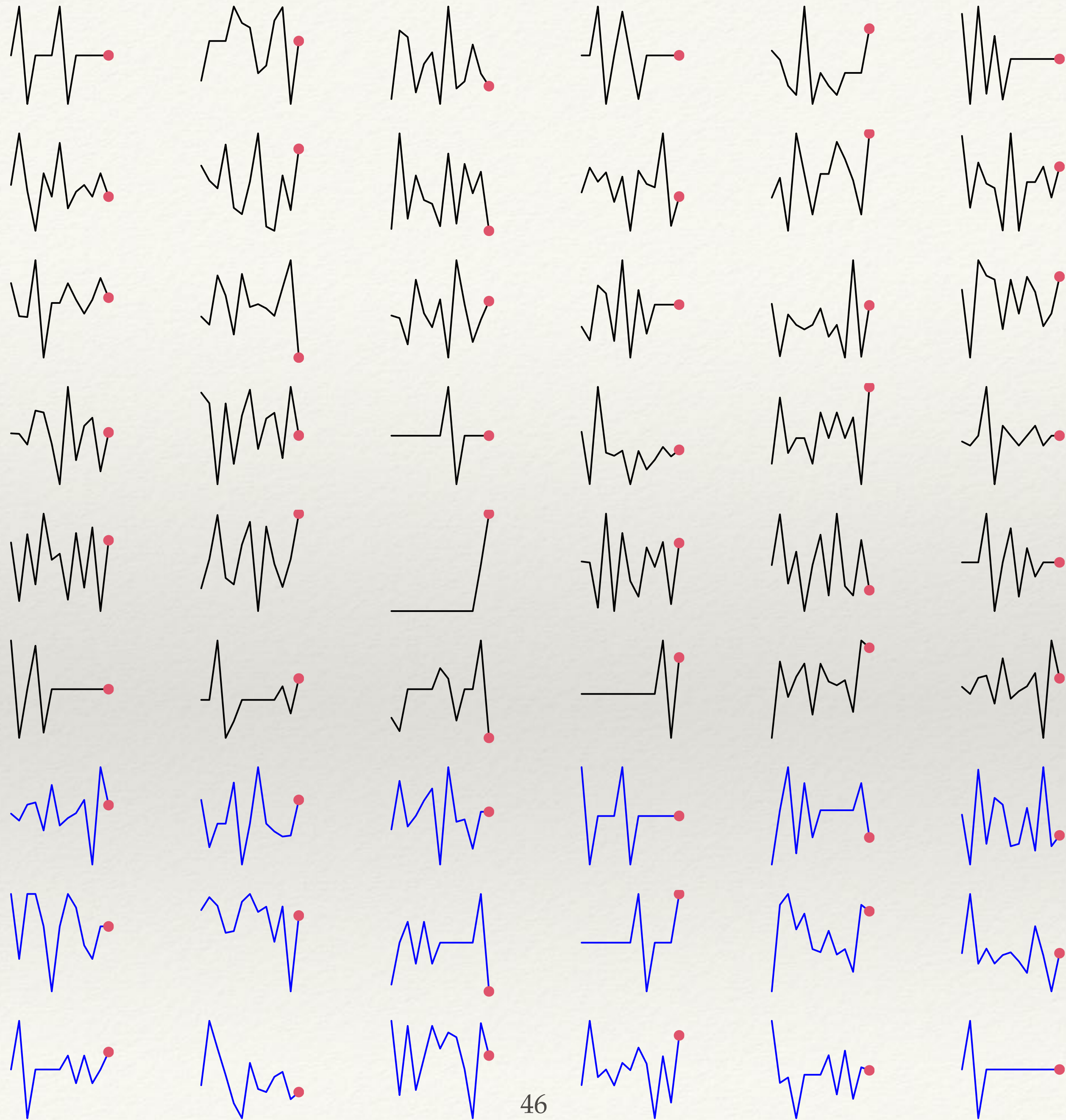


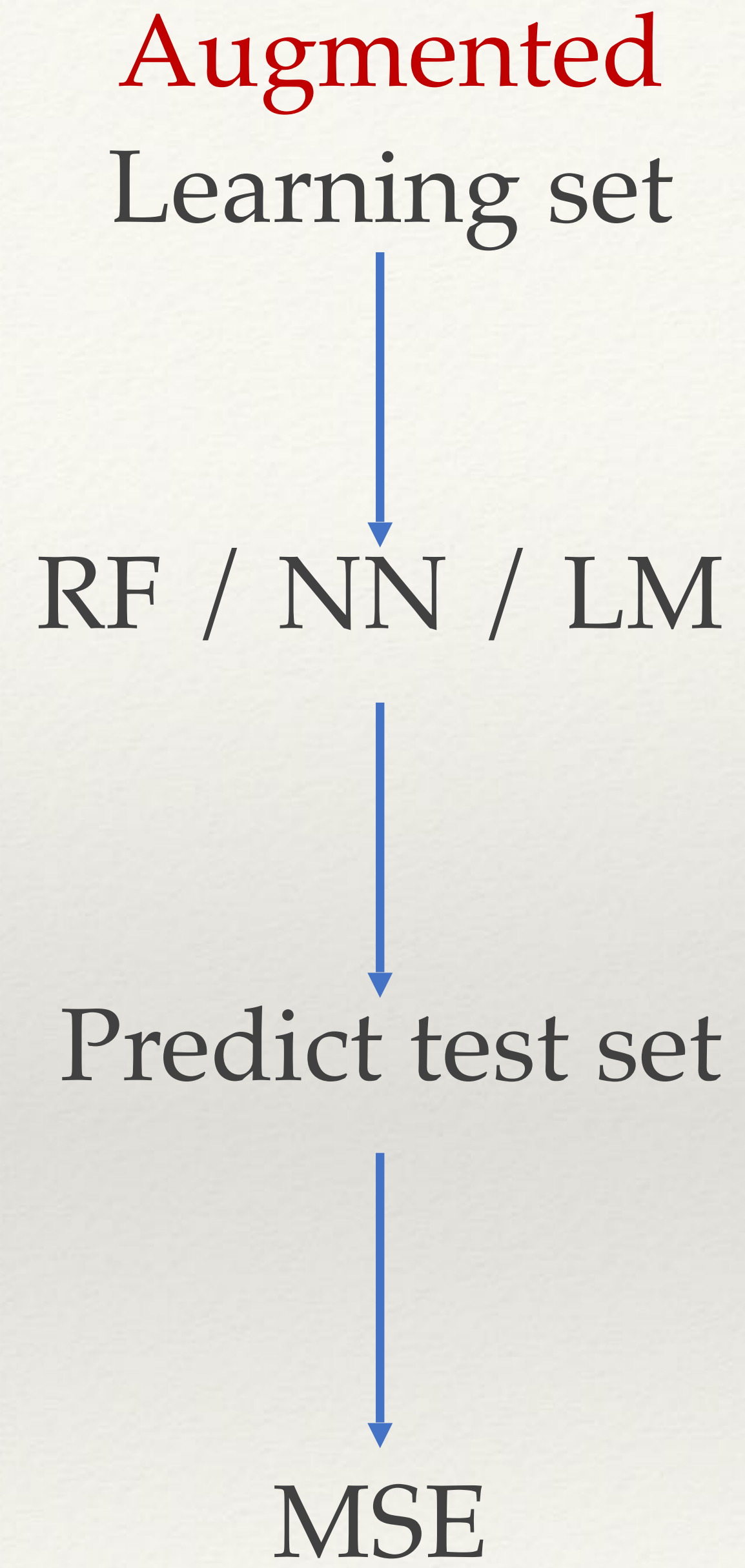
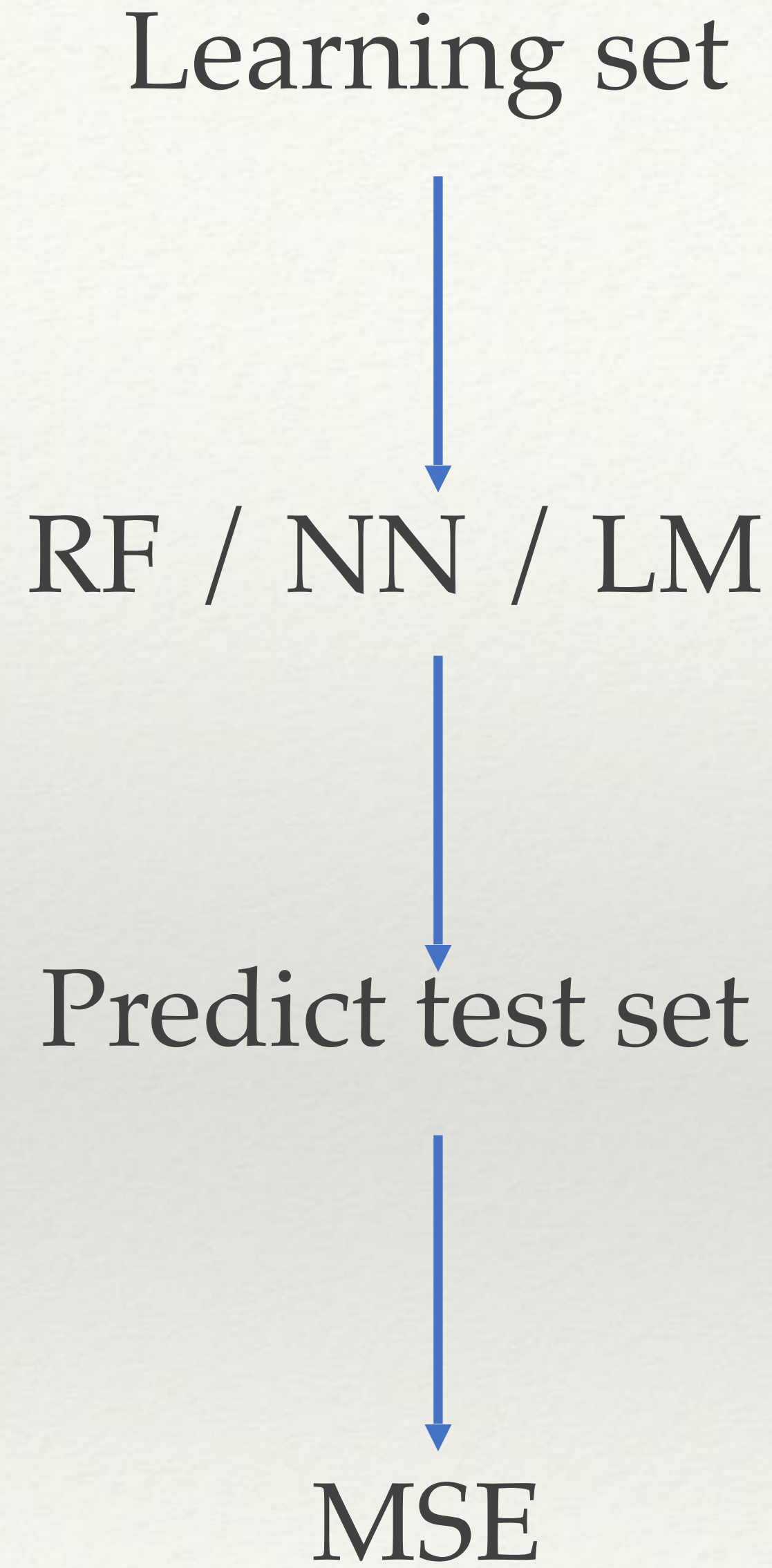
# Basic augmentation for time series: Flipped series

- ❖ **History repeats**
- ❖ **In reverse**
- ❖ **Escalation and de-escalation as mirrors**

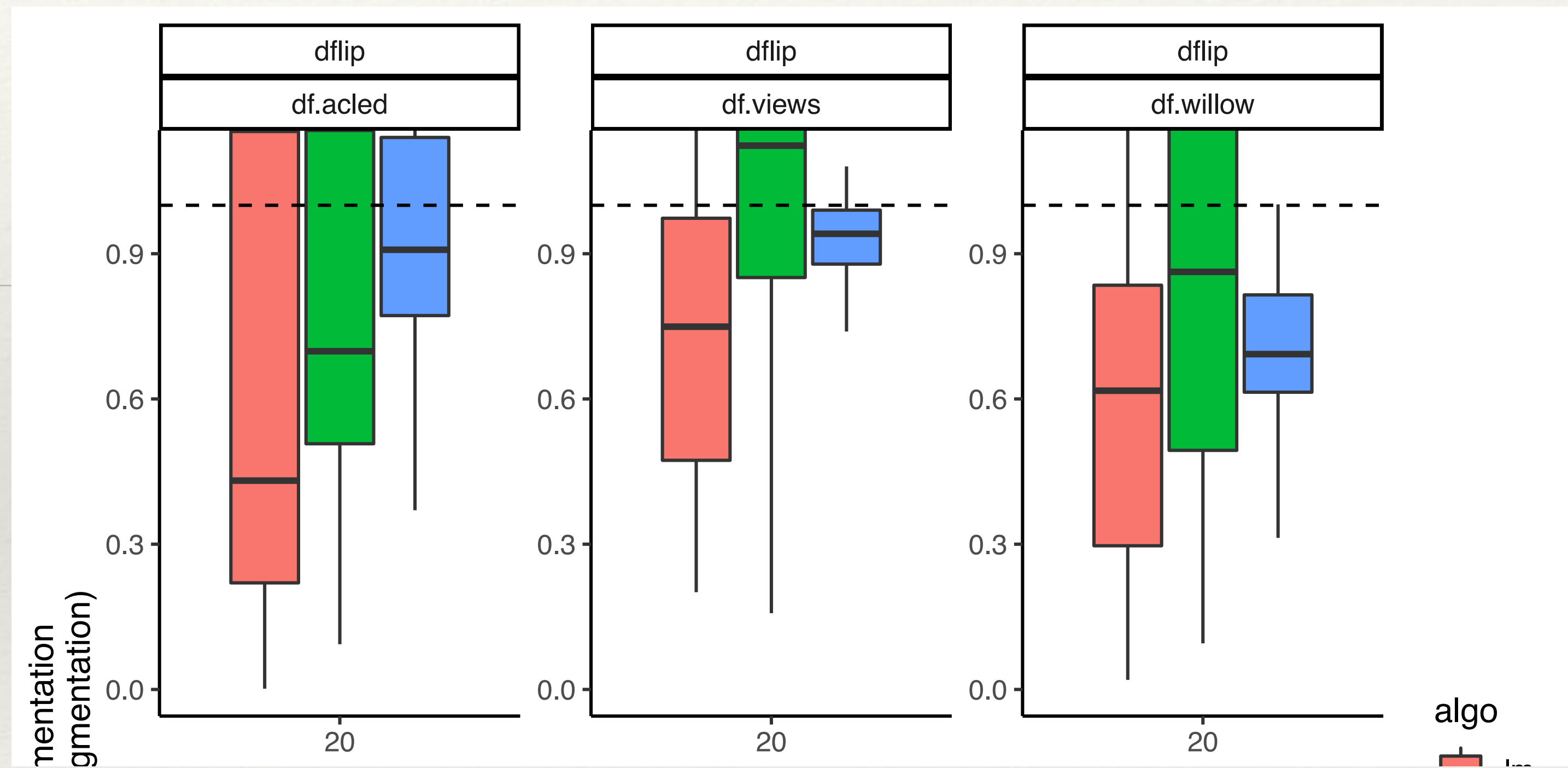


# Augmented Set





# Results: double flip





# How much of an improvement is that really?

MSE with 80 obs



=

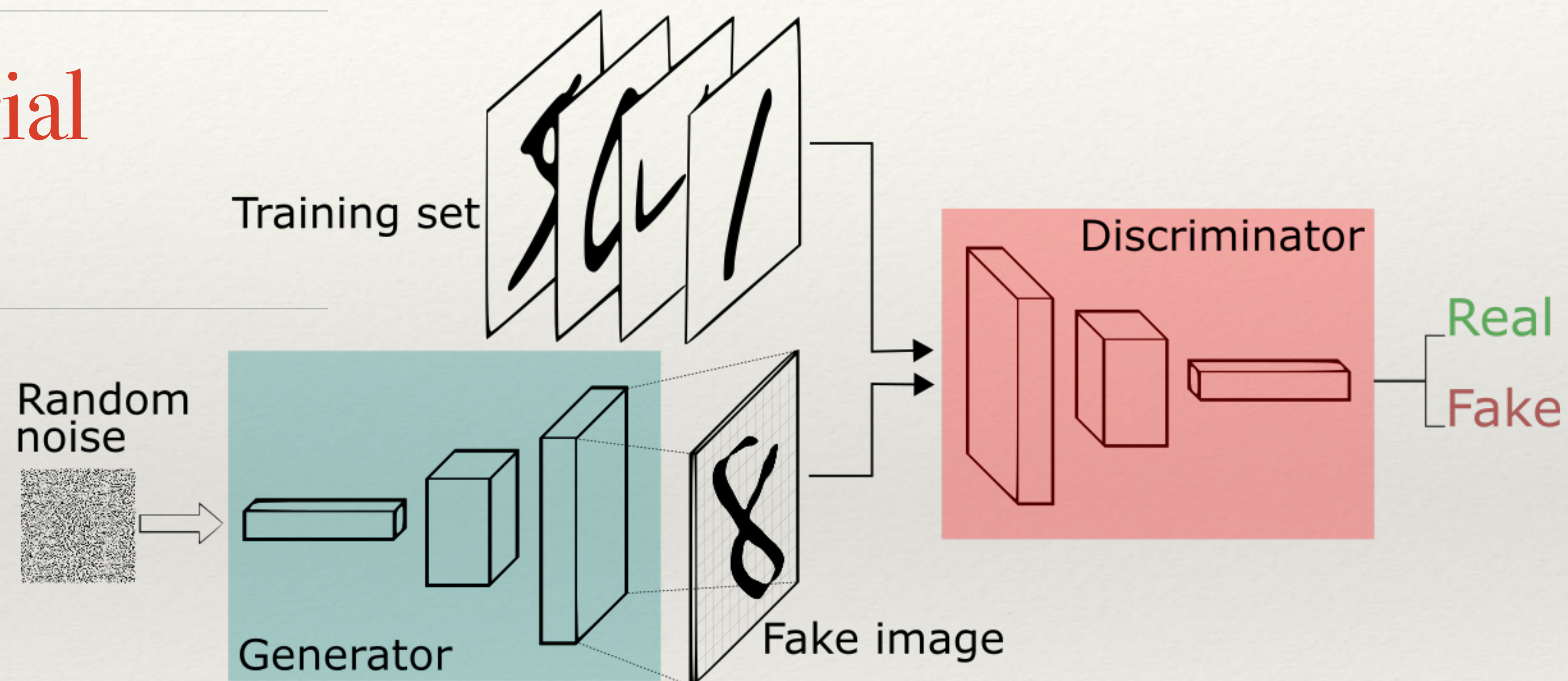
MSE with 100 obs



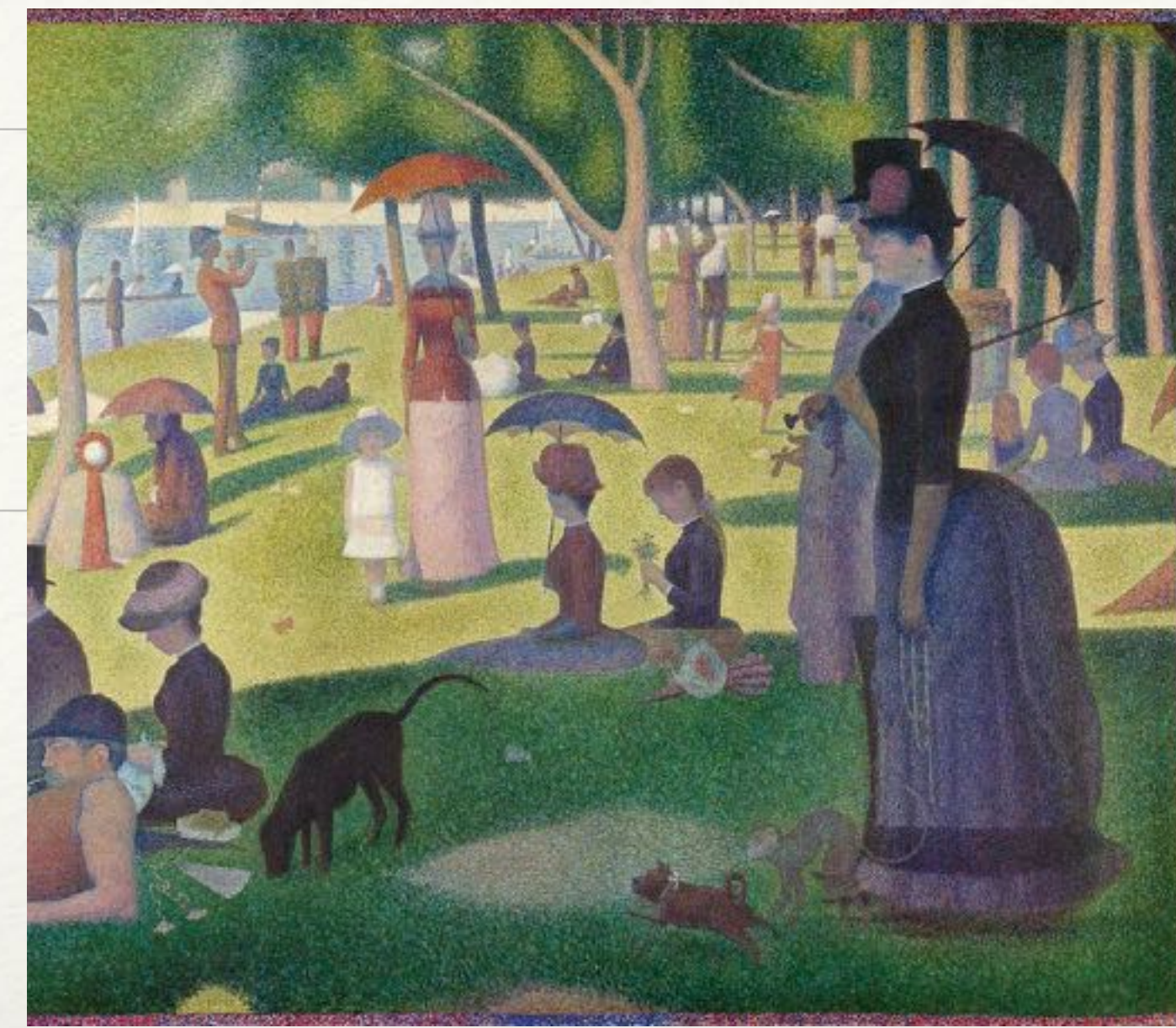
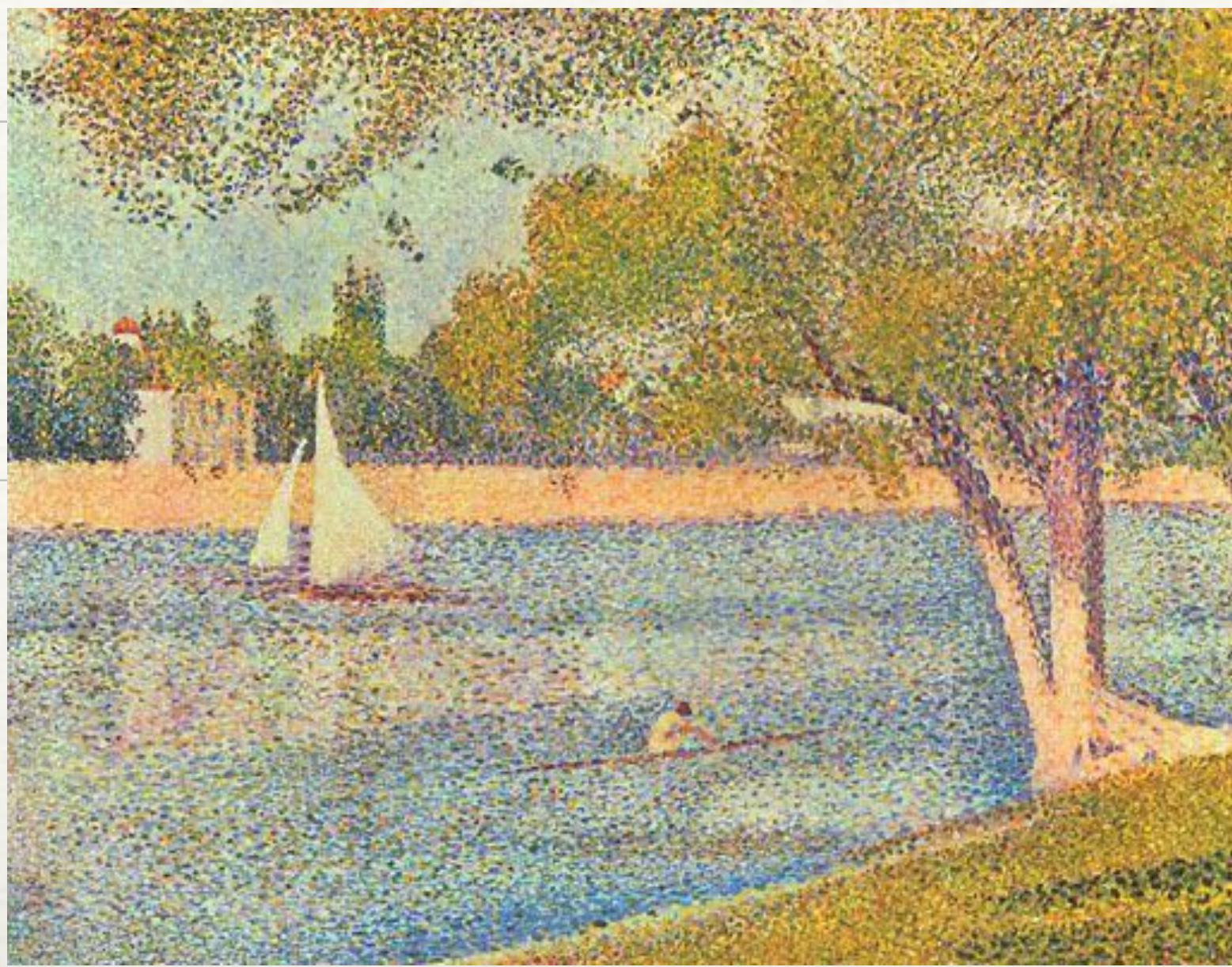
+ double flip Augmentation

# More complex augmentation methods

# Generative adversarial networks



Seurat



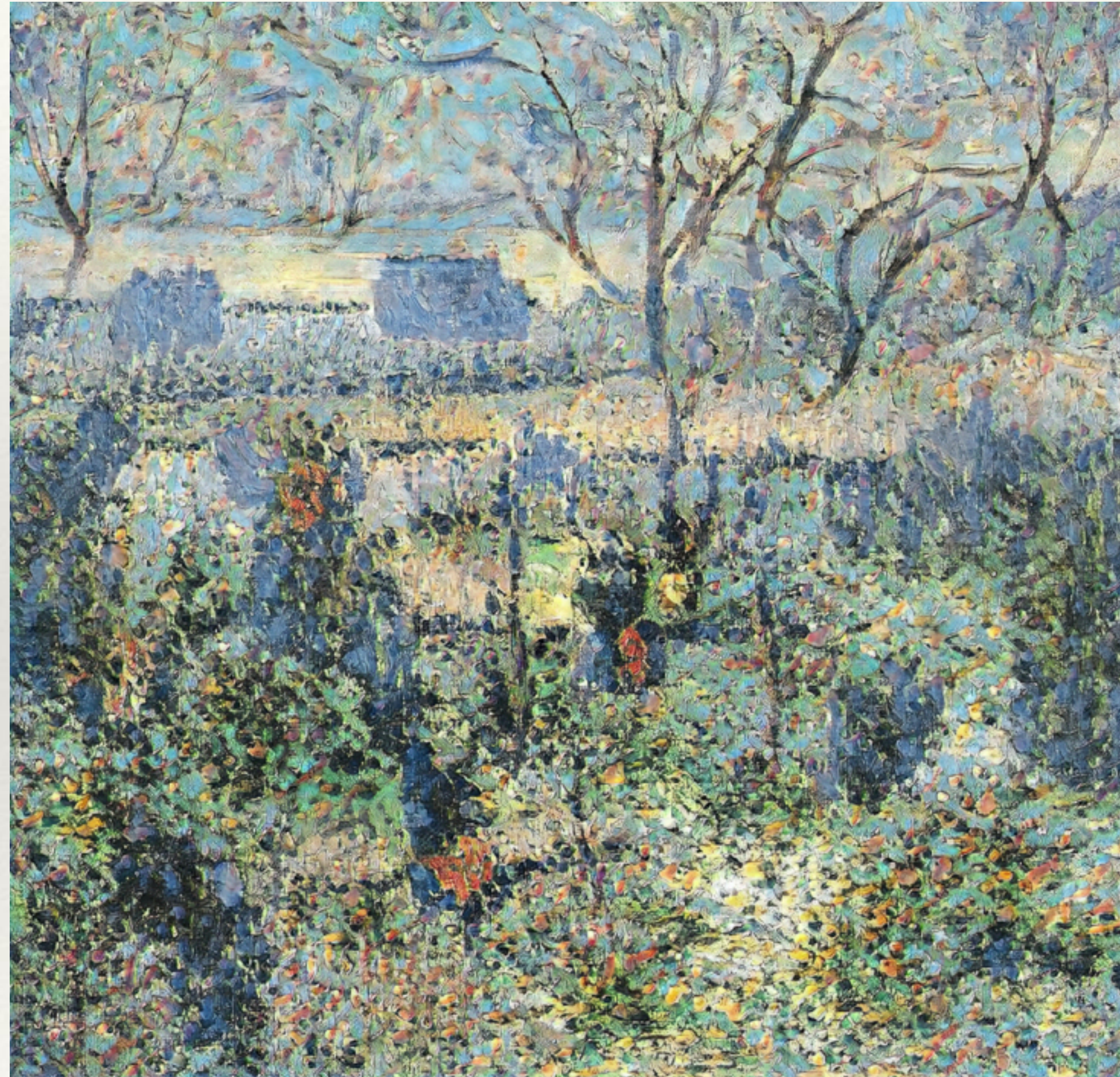
Not  
Seurat



---

# Augmentation: GAN

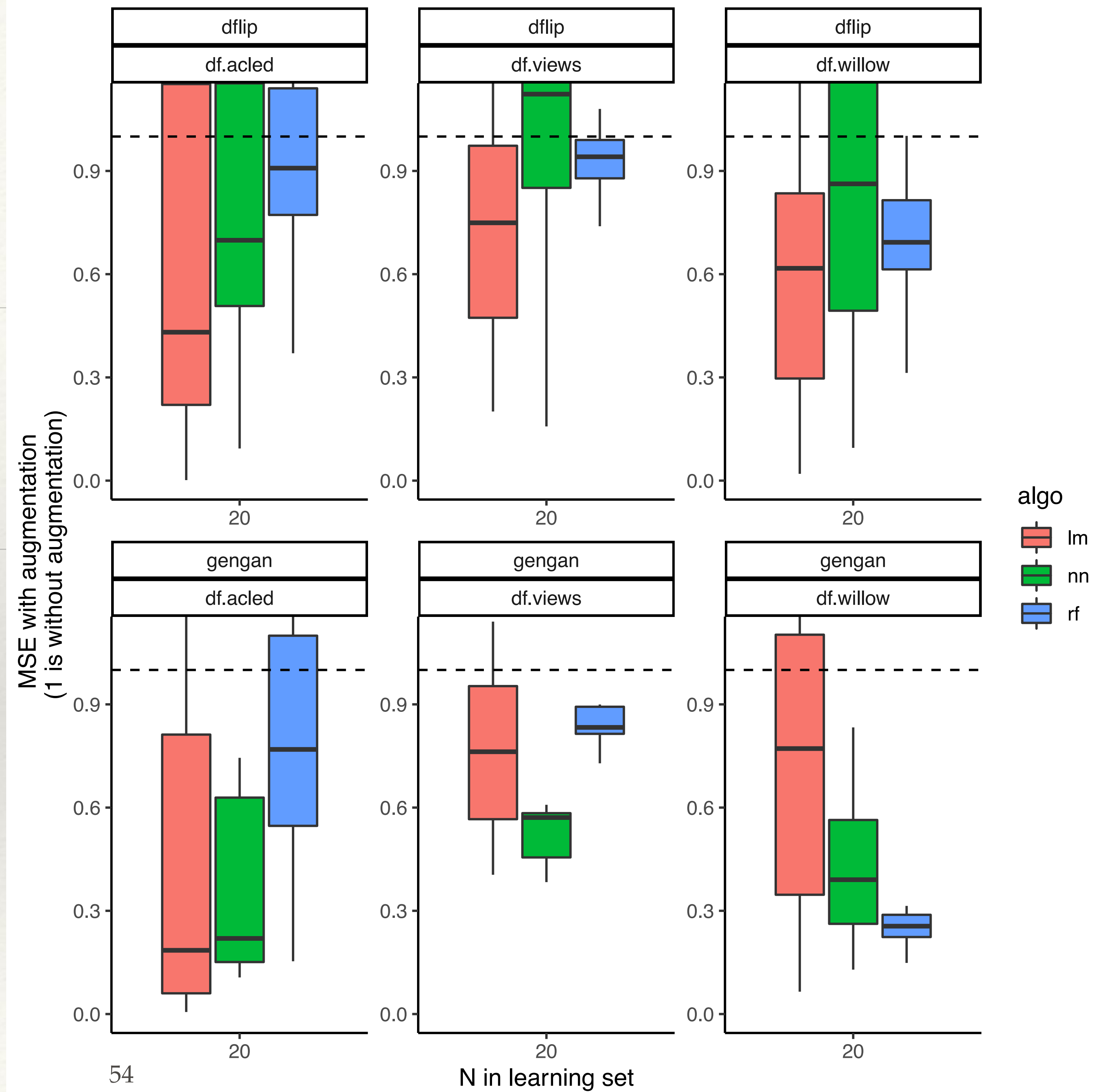
---



Fake Seurat painting

Lorem Ipsum Dolor

# Results: GAN



---

# Todo

---

- ❖ Multi-dimensional data – Panel
- ❖ Other datasets – suggestions?

# Project 4

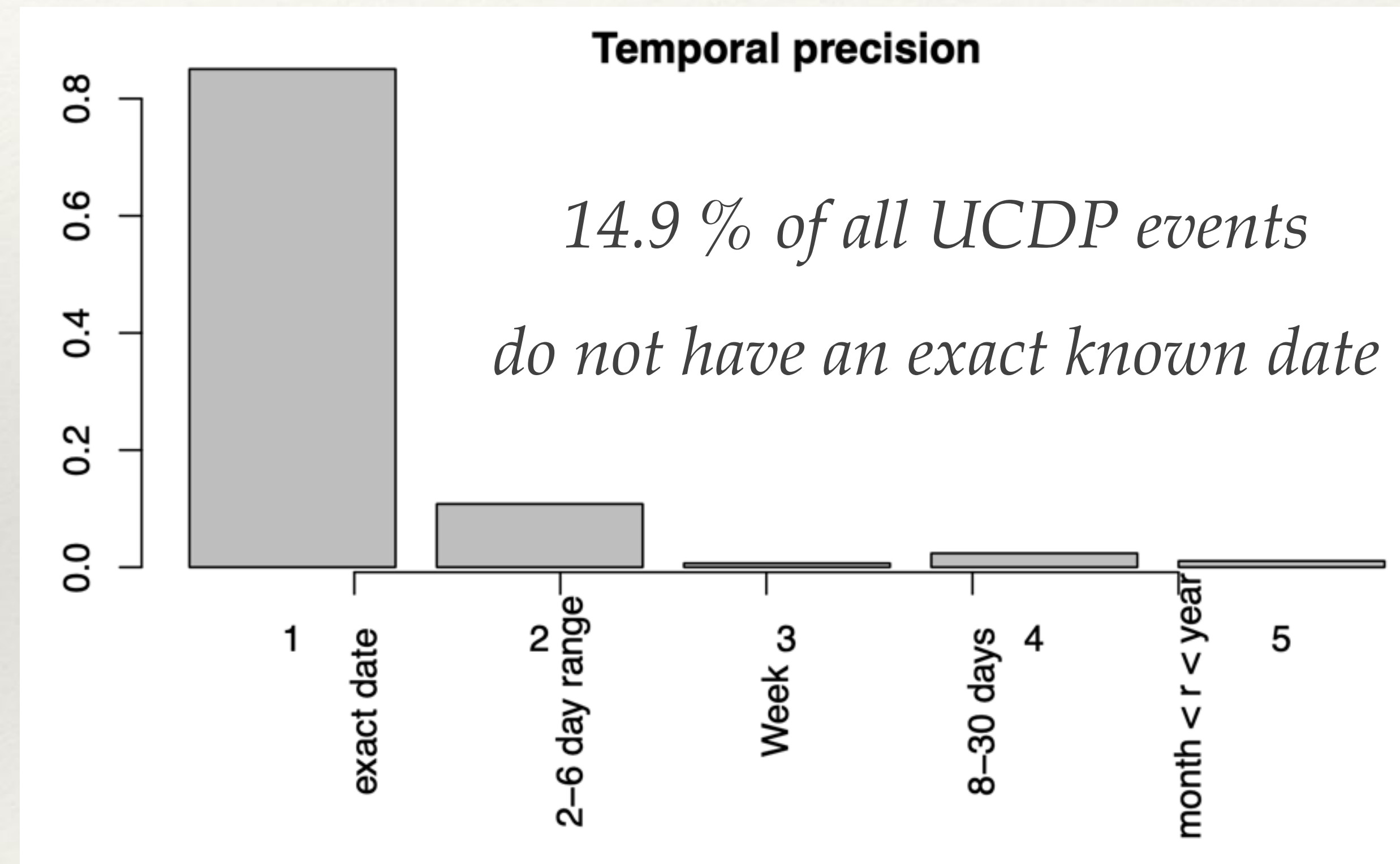
## Reducing Uncertainty In Conflict Events Using Satellite Data

Gareth Lomax & Thomas Chadeaux



# Motivation

- ❖ News based datasets (e.g., UCDP, ACLED) often contain temporal and geographic uncertainty.
- ❖ Uncertainty is compounded by the severity of violence.
- ❖ Datasets do not report quantitative extent of non violent damage.



---

# Satellite Imagery?

---

Contains a wealth of change detection literature.

Medium resolution data is well covered under permissive licences.

(Very) High Resolution data is expensive, unless you know where to look.

# NASA FIRMS

## Fire Information For Resource Management Service

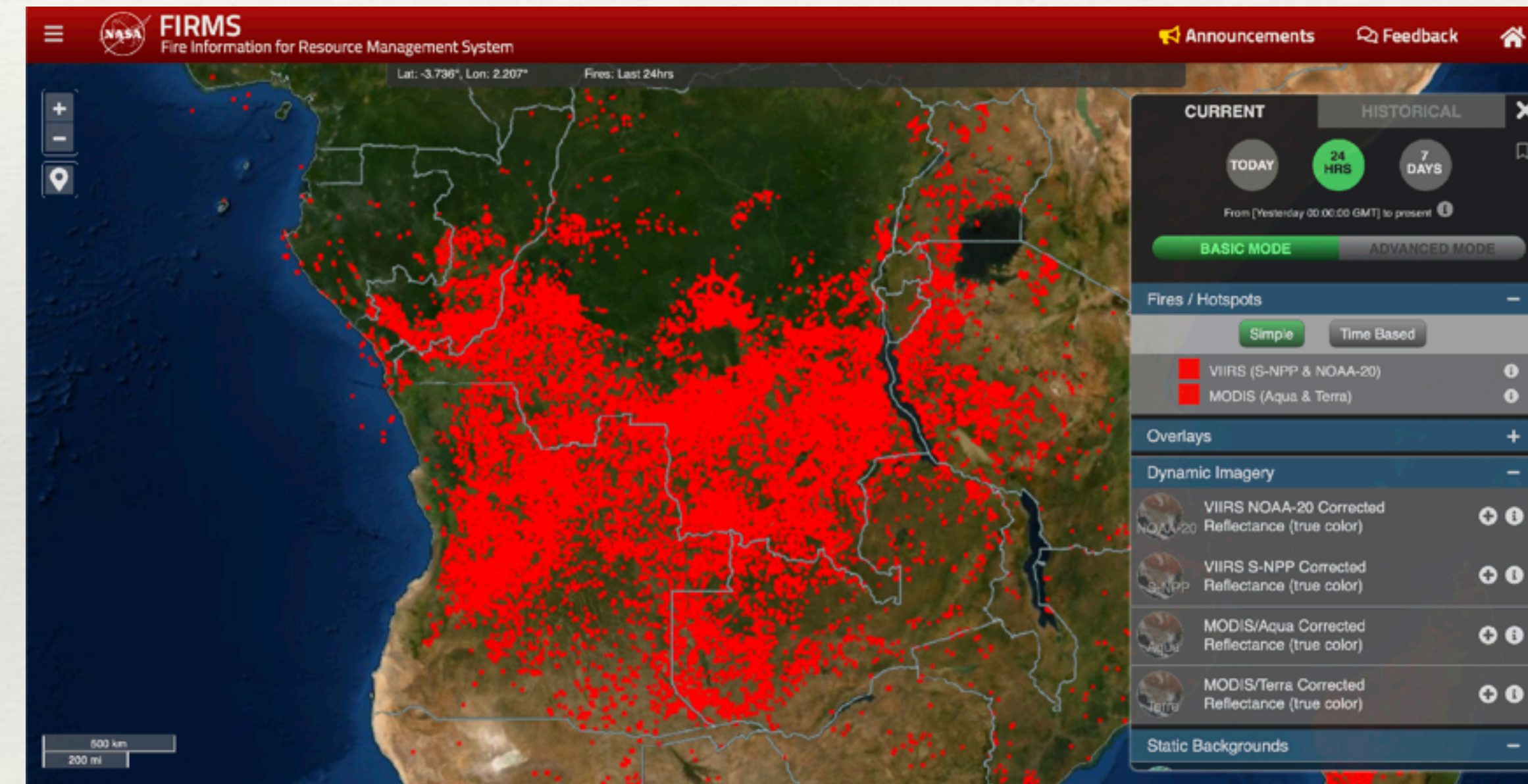
Data from MODIS and VIIRS sensors

MODIS resolution 1km, VIIRS  
resolution 375m

MODIS aboard Terra and Aqua EOS,  
acquisitions at 10:30 a.m. and 10:30 p.m  
MLT

VIIRS aboard Suomi NPP and NOAA-20

Total acquisitions 8 per day.



NASA FIRMS DASHBOARD OVER AFRICA 12/07/2022

---

# Method

---

Use Firms Archive Data to identify statistically unlikely fires to reduce uncertainty in conflict data.

# Myanmar

- ❖ Ethnic cleansing committed by the Tatmadaw (Military) from 2017 - 2018 lead to large scale displacement and destruction.
- ❖ Approximately 900,000 Rohingya displaced.
- ❖ In Rakhine State 30.8% of events have no exact date. 84% of events have no precise location.
- ❖ Open Source Projects e.g Ocelli Project Seek to reduce this uncertainty.



---

# Ocelli Project

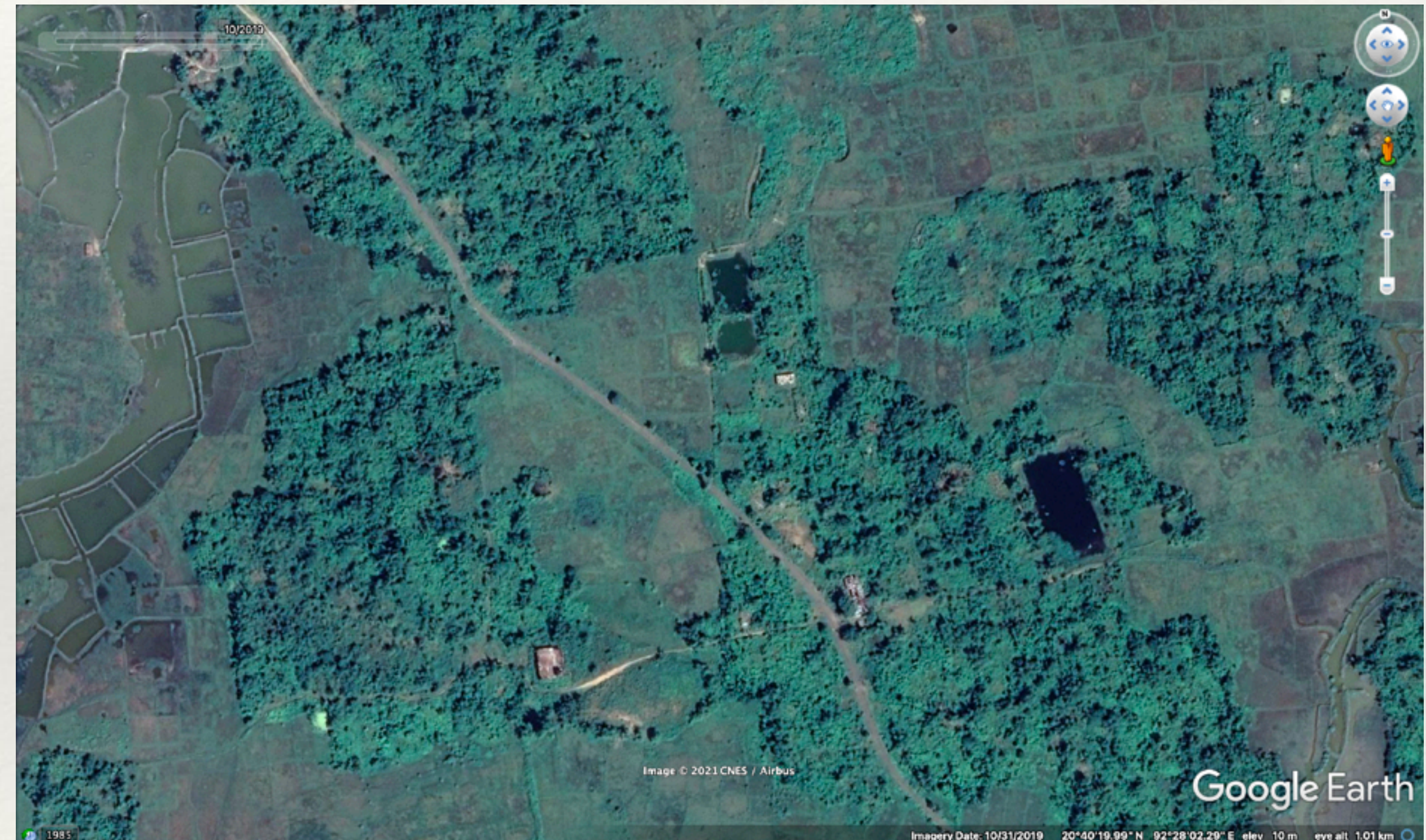
---

- ❖ Dataset of 613 (620) Rohingya Settlements destroyed between 2017 and 2018.
- ❖ Based on open source very high resolution satellite images (google earth) to identify village burnings manually.
- ❖ Mixture of satellite and aerial imagery.

# Anauk Myinhlut and Al Le Than Kyaw area



December 27, 2016



October 31, 2019



Image © 2021 Maxar Technologies

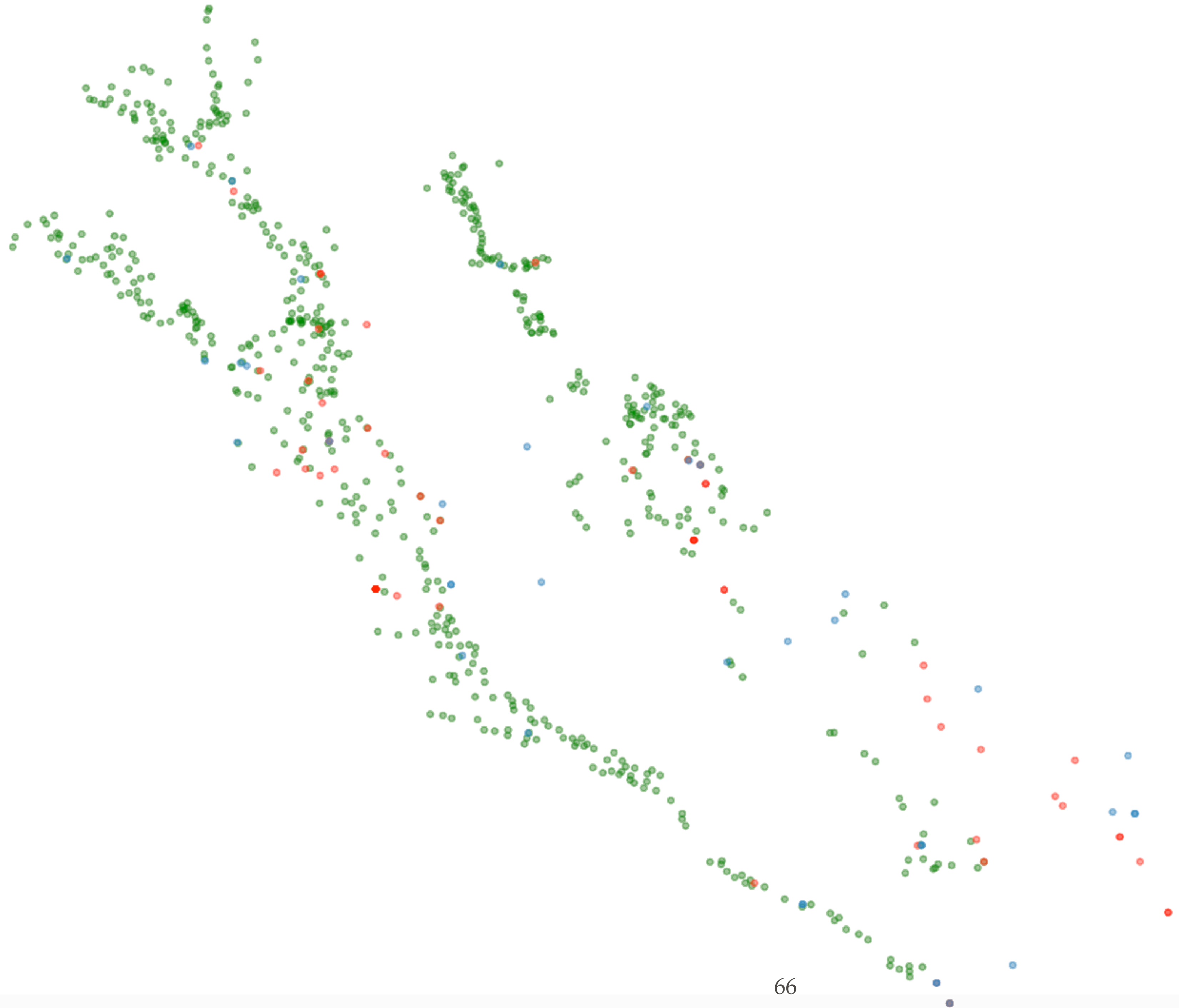
Google Earth





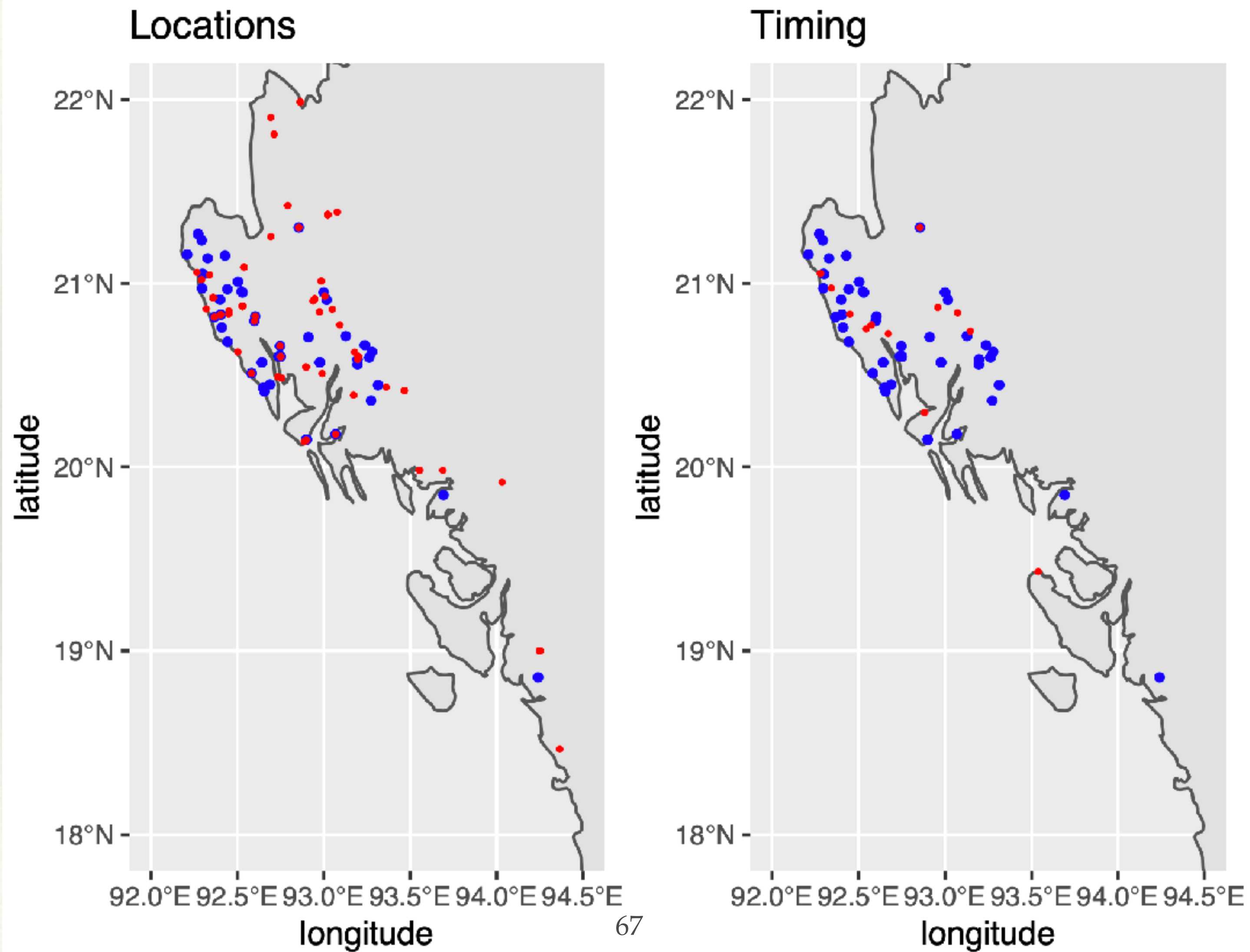
Image © 2021 Maxar Technologies

Google Earth

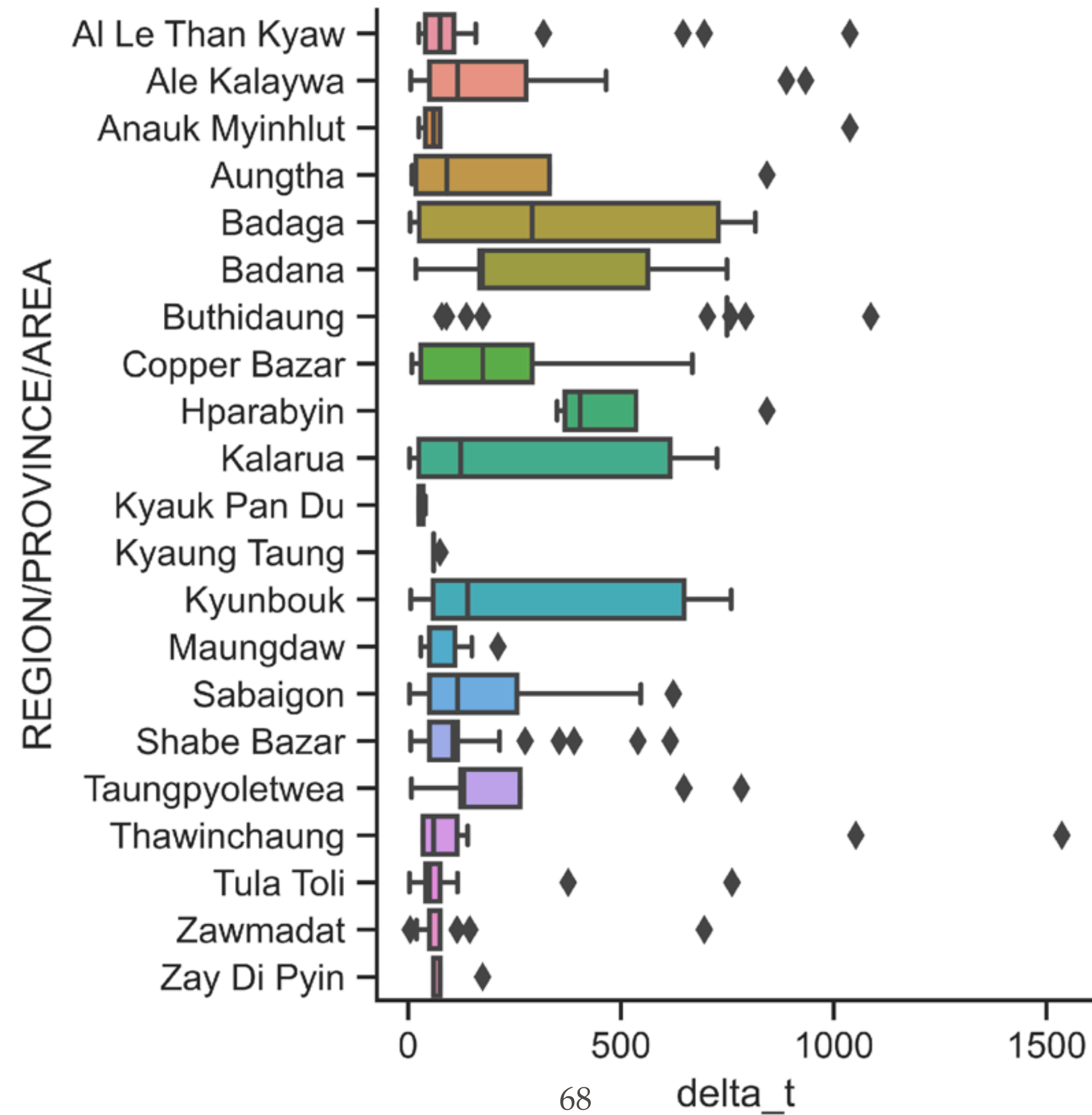


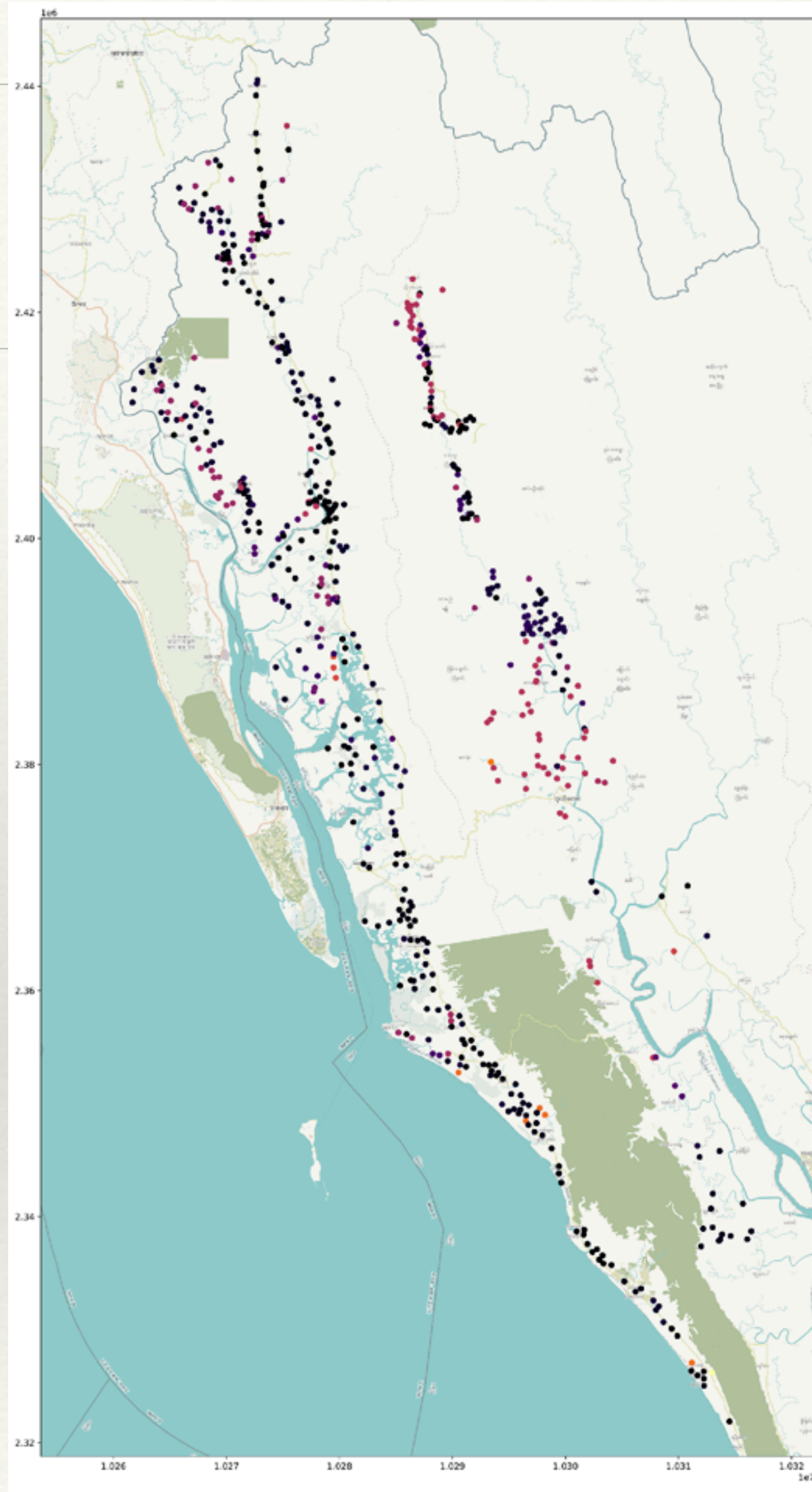
- ❖ **Green: Ocelli**
- ❖ **Red: ACLED**
- ❖ **Blue: UCDP**

# Uncertainty in UCDP data (Rakhine)

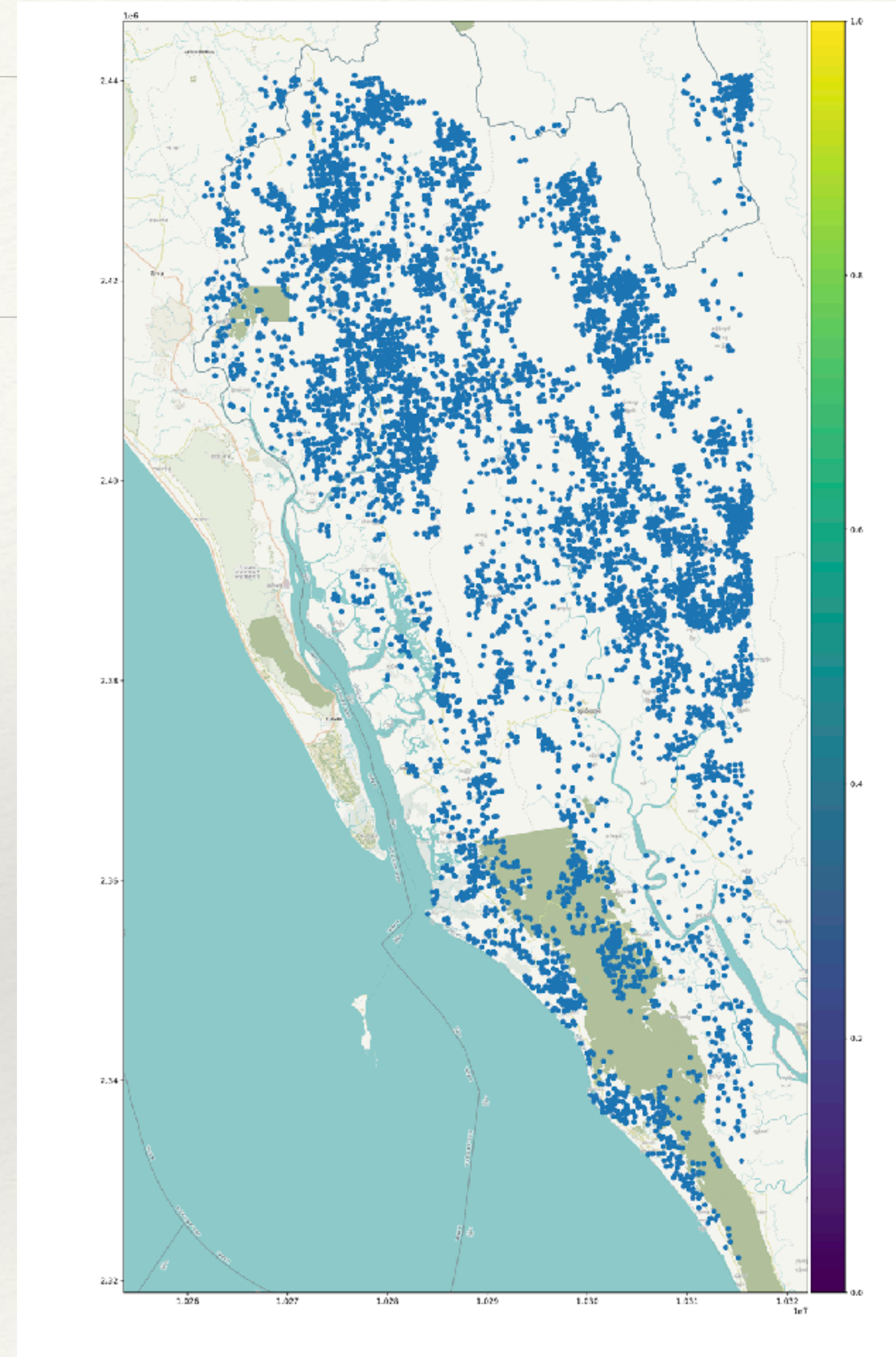


# Uncertainty in Ocelli data (Rakhine)





Event data (Ocelli)



Fire data (FIRMS)

---

# Method

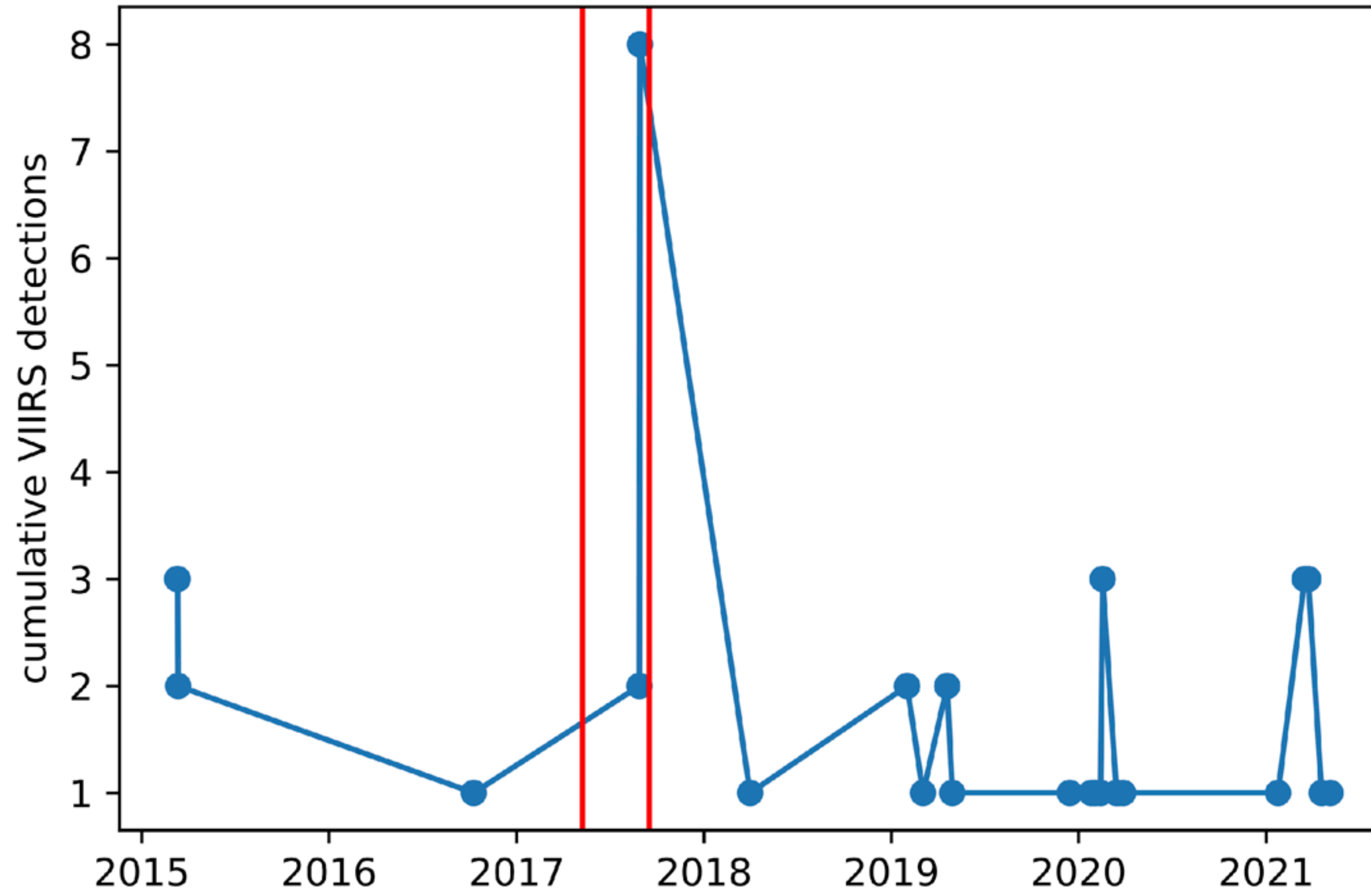
---

Fire data is highly seasonal due to regular wildfires and agricultural burnings.

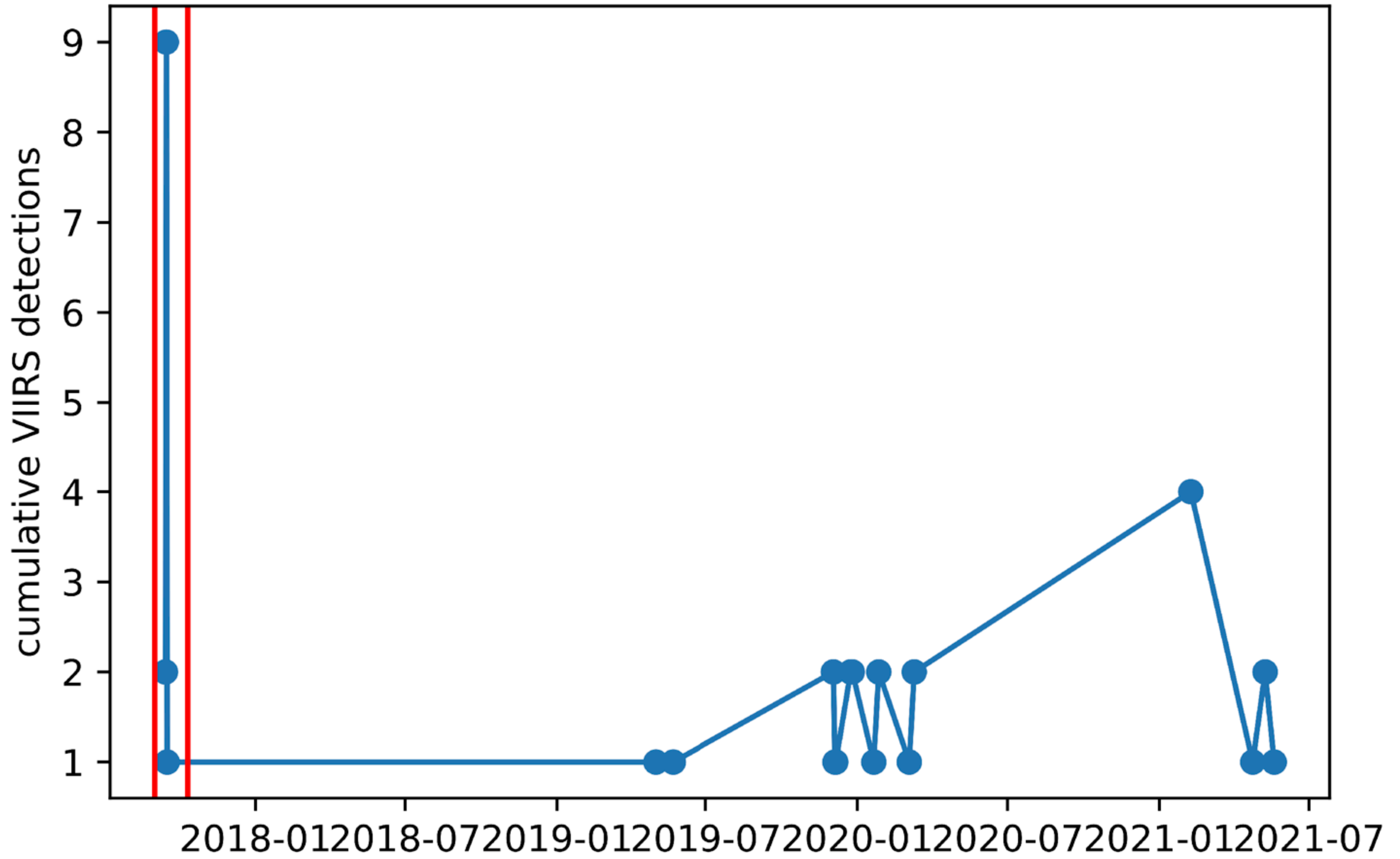
Extract localised burnings in radius about towns and compare to local minima.

Anomaly detection methods compared: Naive KDE's, ARIMAX residuals, Seasonally adjusted temporal KDE's.

# L5-013

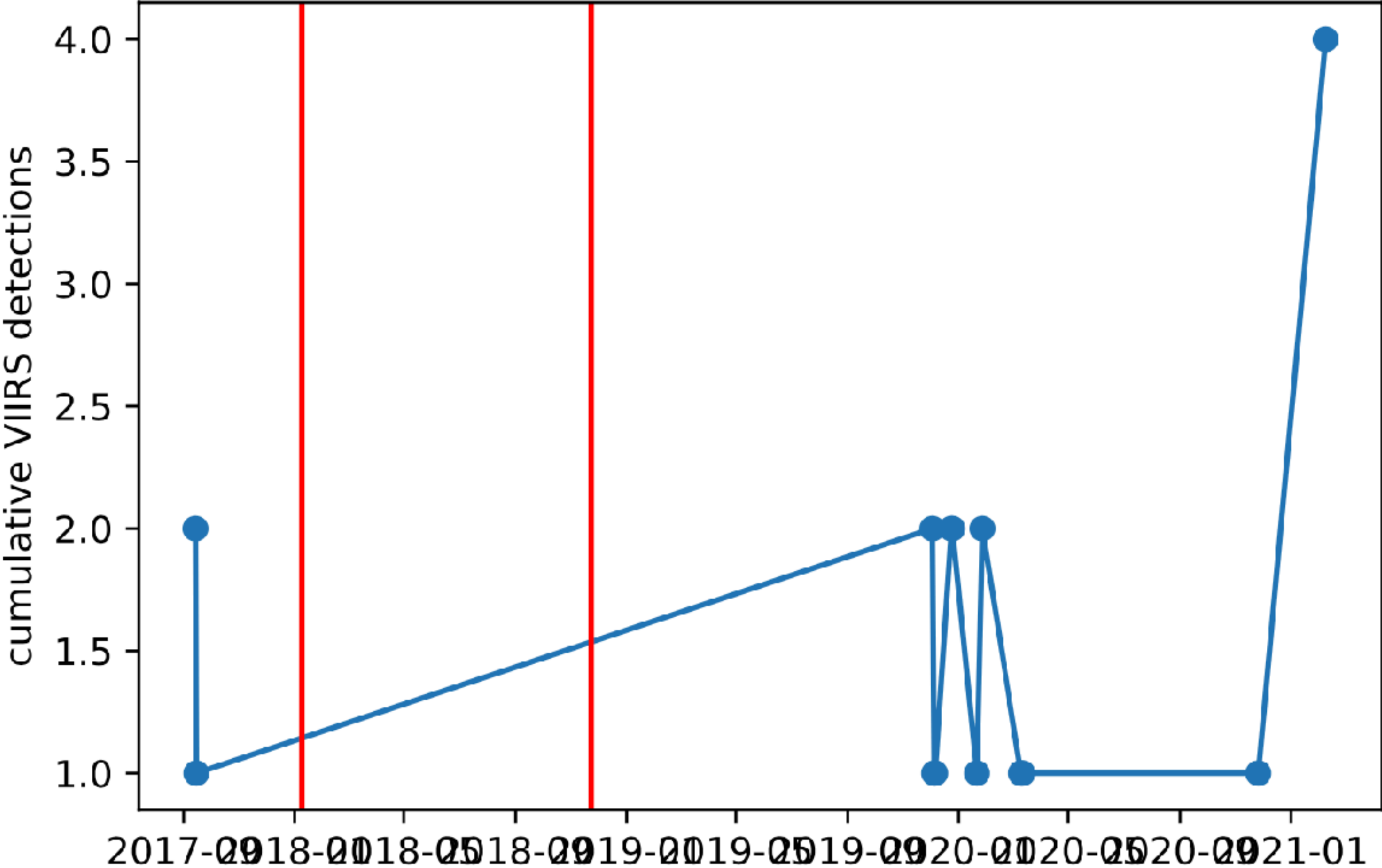


# Q6-008





Q6-003

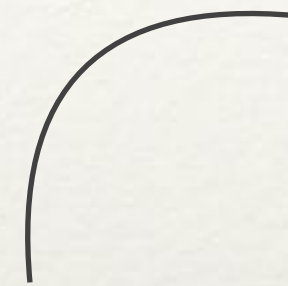


---

# Overall performance

---

Seasonal KDE accuracy:



- Pick a town-day at random. 13% of the time it will fall within an Ocelli window. I.e., a 13% chance of correctly “timing” an event.
- Using the anomalous fires, the same happens 41% of the time
- If we only consider only areas where burning actually took place, then 63%.

---

# Why it matters

---

- ❖ News-based reporting often lacks precision. Also makes real-time reporting impossible
- ❖ Solution: satellite data. BUT expensive
- ❖ -> need to know where to look
  - ❖

---

# Fires: pros and cons

---

- ❖ Pros:

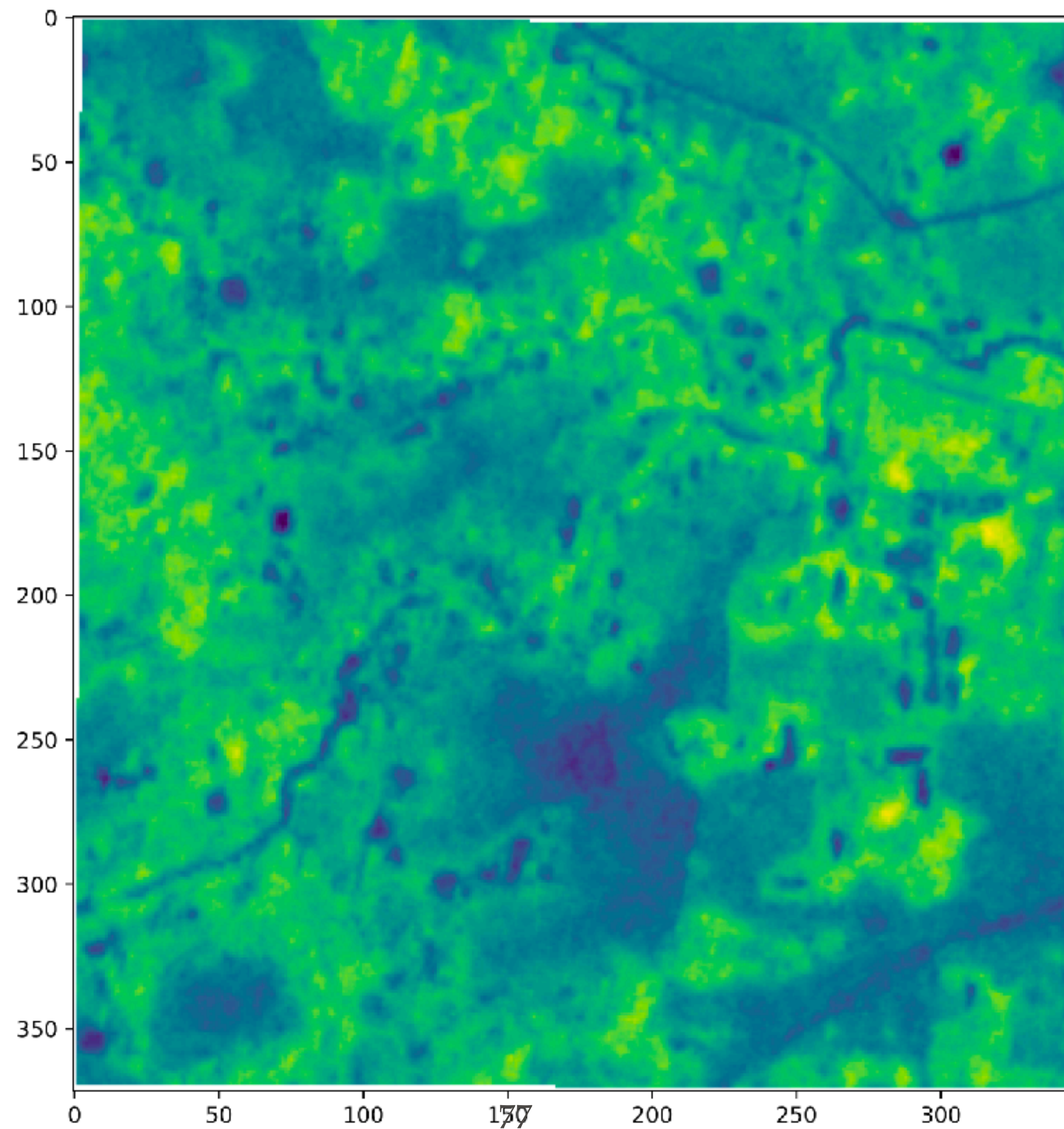
- ❖ go back 20 years
- ❖ cheap
- ❖ instantaneous (3 hrs)
- ❖ Allow to estimate news-based datasets' underreporting factor



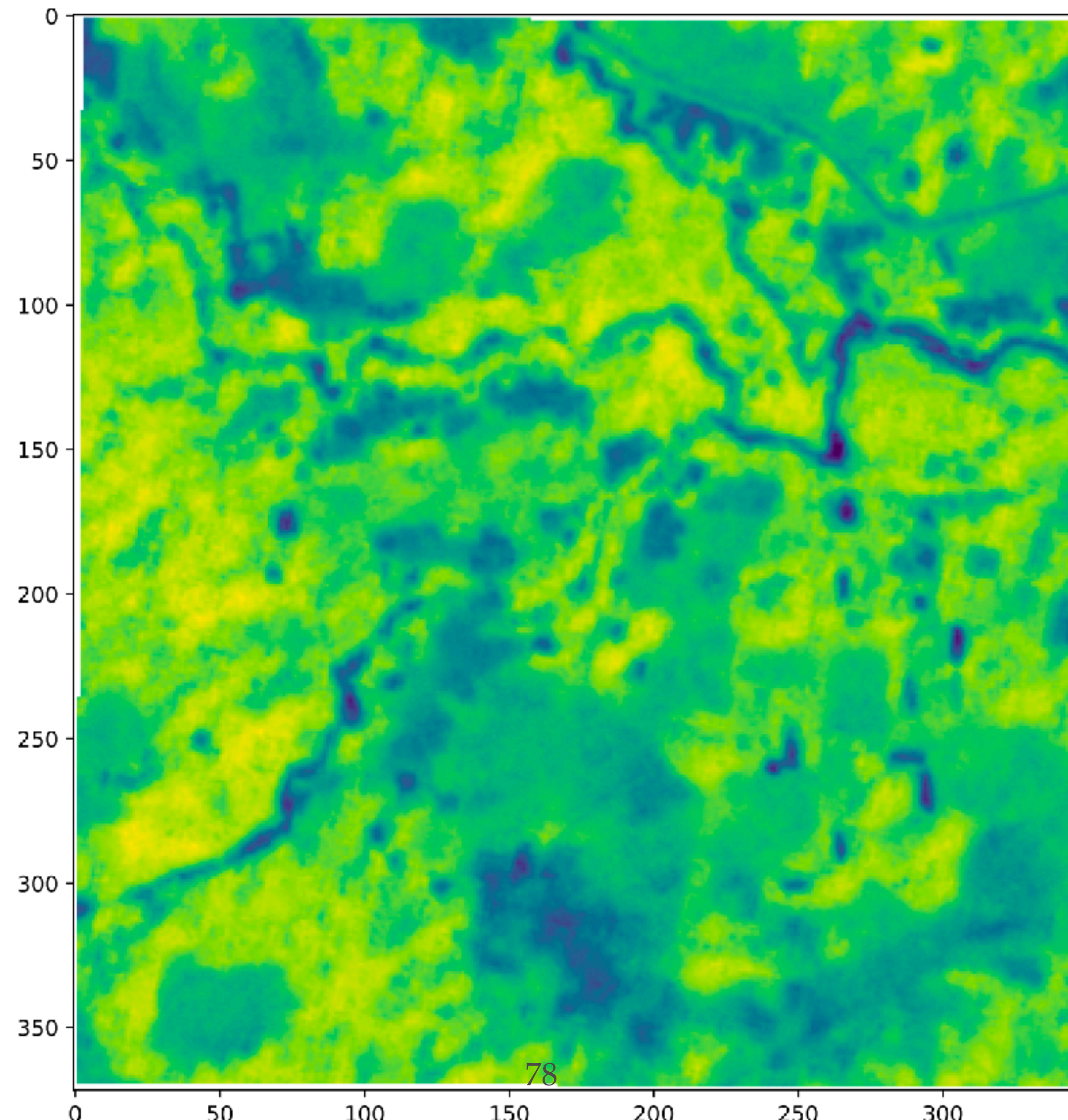
- ❖ Cons:

- ❖ Limited to events that cause fires
- ❖ Low accuracy for now

# Next steps: Reflectance



# Next steps: Reflectance



---

# Next steps

---

- ❖ Assign fires to closest Ocelli event, as opposed to all within a radius -> accuracy will improve
- ❖ Remove false positives in uninhabited locations
- ❖ Earth mover distance
- ❖ Improve seasonal KDE

# Project 5

*Dynamic Synthetic Controls: Accounting for Varying Speeds in Comparative Case Studies*

*Jian Cao & Thomas Chadeaux*



## Research Question

One of the most popular comparative analysis methods, synthetic control, **cannot account for potential different speeds** in time series.

# Research Question

One of the most popular comparative analysis methods, synthetic control, **cannot account for potential different speeds** in time series.

**Can we design a method that can solve this problem?**

## 1. Limitation of Synthetic Control

1. Limitation of Synthetic Control
2. Two-Fold Dynamic Time Warping

1. Limitation of Synthetic Control
2. Two-Fold Dynamic Time Warping
3. Simulation Study

1. Limitation of Synthetic Control
2. Two-Fold Dynamic Time Warping
3. Simulation Study
4. Re-run Abadie's Analyses

1. Limitation of Synthetic Control
2. Two-Fold Dynamic Time Warping
3. Simulation Study
4. Re-run Abadie's Analyses
5. Conclusion

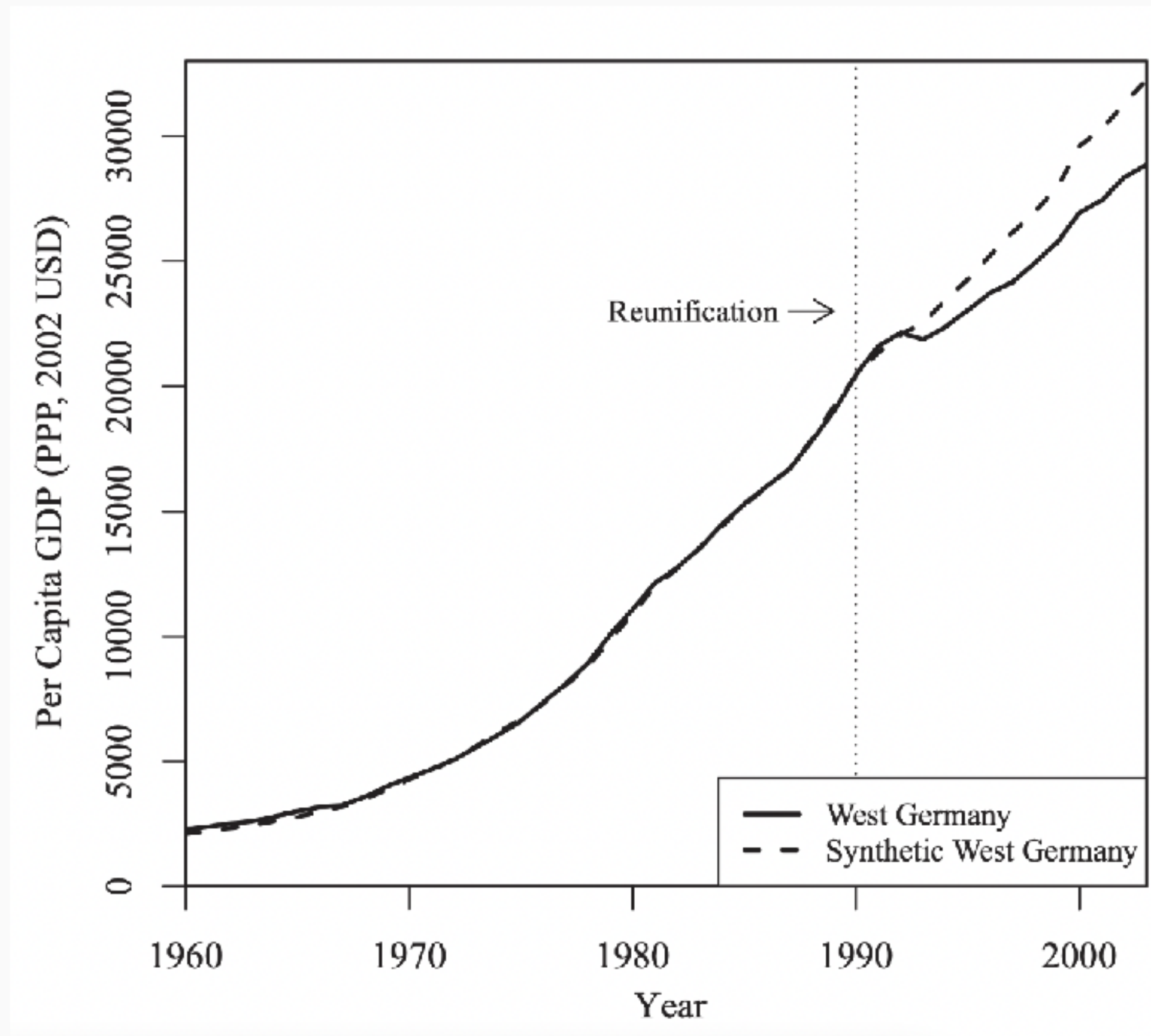
# Limitation of Synthetic Control

---



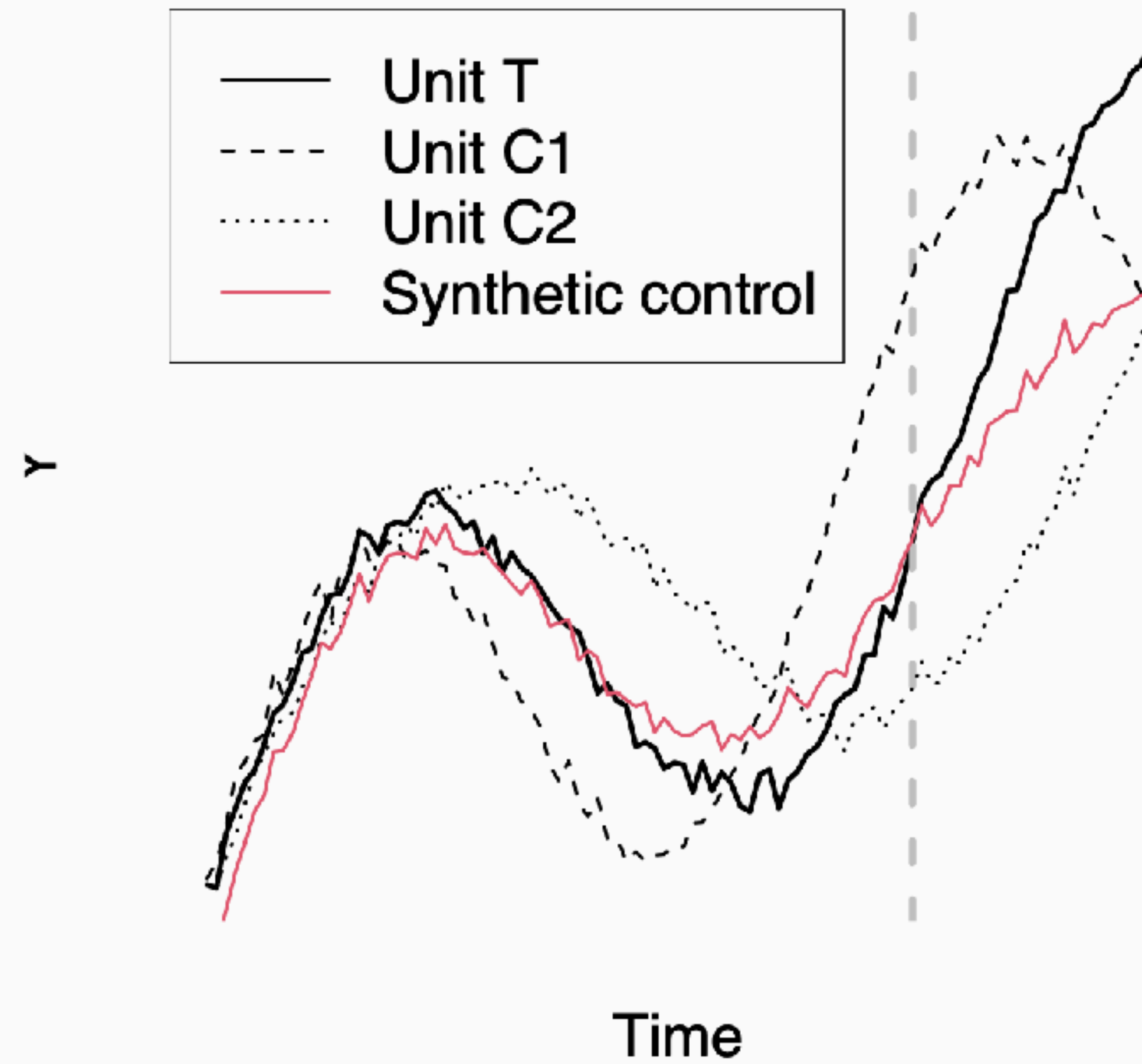
# Synthetic Control

## The Economic Cost of the 1990 German Reunification



# Synthetic Control (Different Speed)

## Illustrative Example



# Two-Fold Dynamic Time Warping

---

## What is TFDTW?

- Two-Fold Dynamic Time Warping (TFDTW) is a new DTW method that is designed specifically for comparative case studies, and helps them account for different speeds in time series.

## What is TFDTW?

- Two-Fold Dynamic Time Warping (TFDTW) is a new DTW method that is designed specifically for comparative case studies, and helps them account for different speeds in time series.
- It is a pre-processing method.

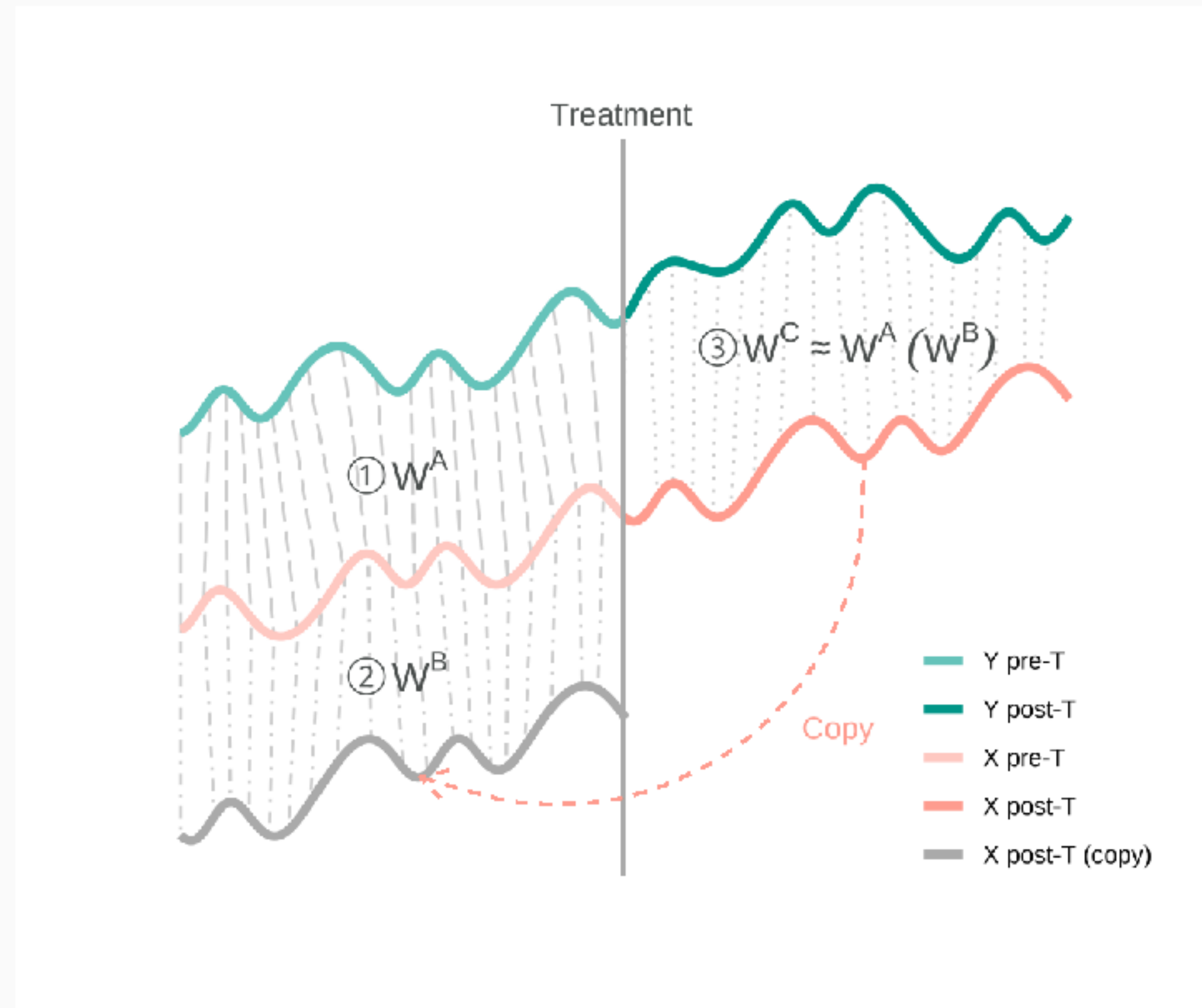
## What is TFDTW?

- Two-Fold Dynamic Time Warping (TFDTW) is a new DTW method that is designed specifically for comparative case studies, and helps them account for different speeds in time series.
- It is a pre-processing method.
- It warps time series to match the speed of the target time series.

## What is TFDTW?

- Two-Fold Dynamic Time Warping (TFDTW) is a new DTW method that is designed specifically for comparative case studies, and helps them account for different speeds in time series.
- It is a pre-processing method.
- It warps time series to match the speed of the target time series.
- Any comparative case study methods, e.g. synthetic control, can be used after TFDTW as differences in speeds have been removed.

## Two-Fold Dynamic Time Warping





# Simulation Study

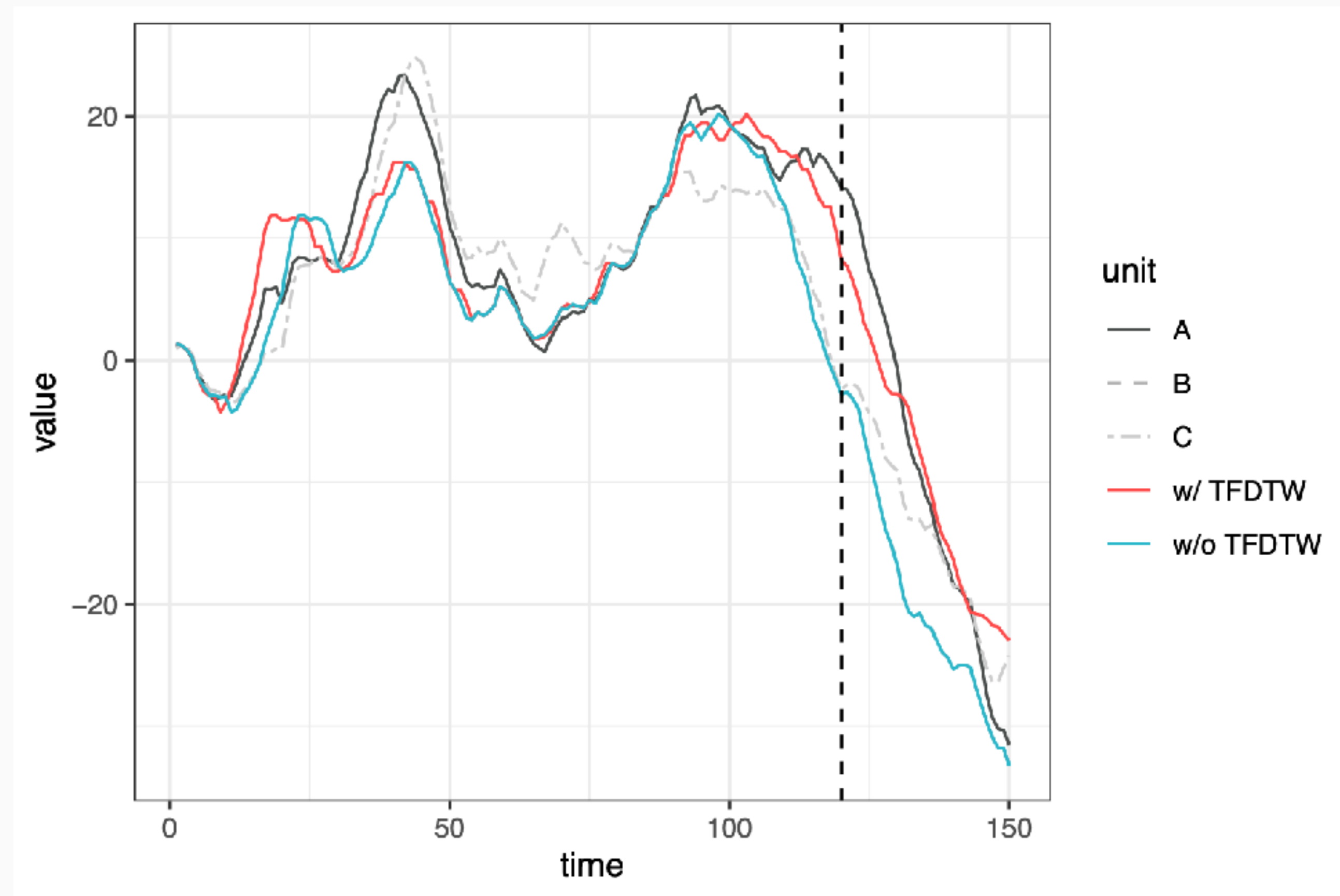
---

- Data generation process (DGP):

$$y_{i,t} = \sigma_t + \sum_{l=1}^p (\beta_{i,t,l} y_{i,t-l} + \phi_{i,t,l} x_{i,t-l}) + \varepsilon_{i,t}$$

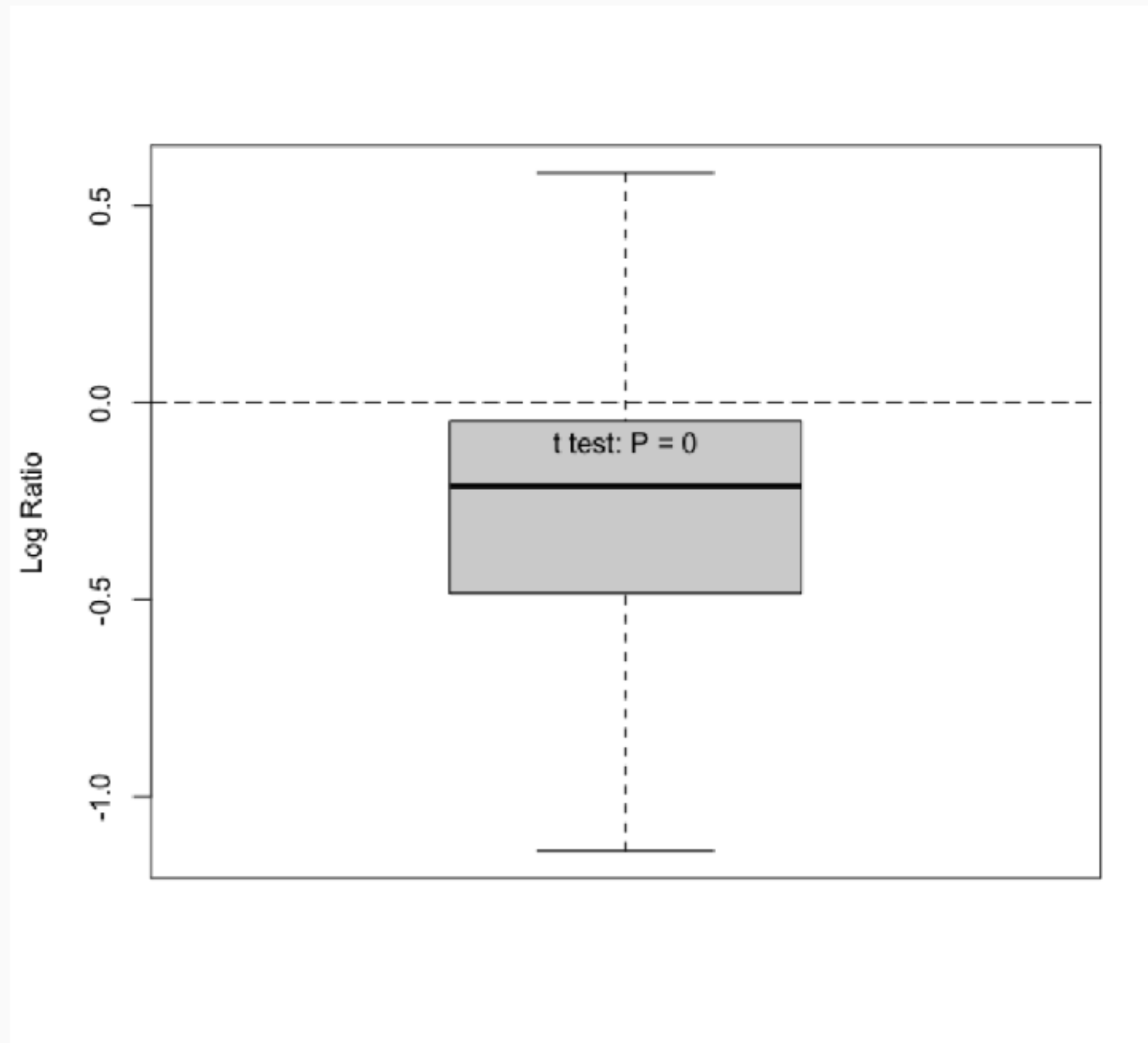
- Simulate 1,000 data sets

## Simulation Example



# Simulation Study

Simulation Result: 1000 Data Sets



$$\text{Log Ratio} = \log(MSE_{w/TFDTW} / MSE_{w/oTFDTW})$$

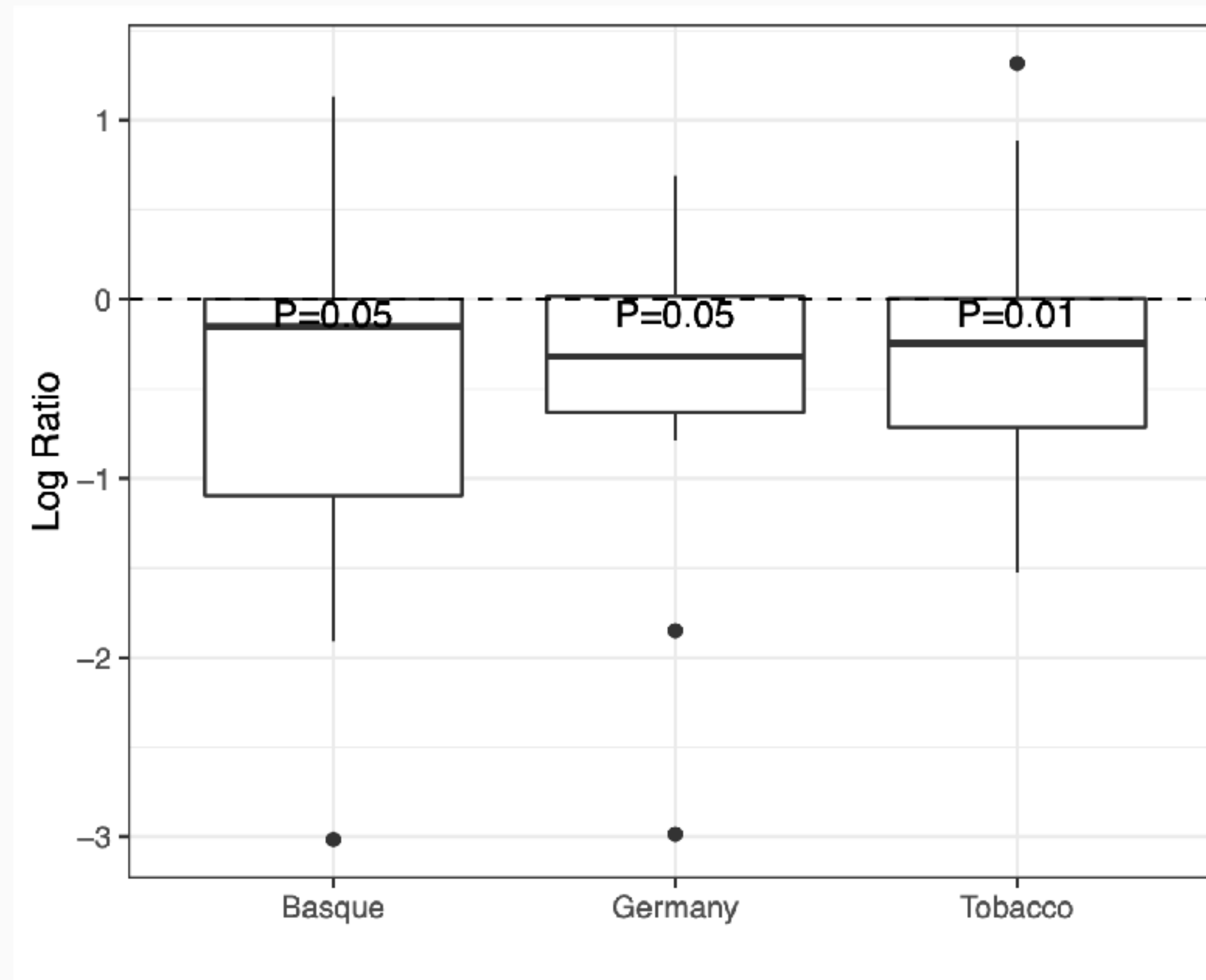
# Re-run Abadie's Analyses

---

- German Reunification - 1990
- California Tobacco Law - 1989
- Terrorism in Basque Country, Spain - 1970

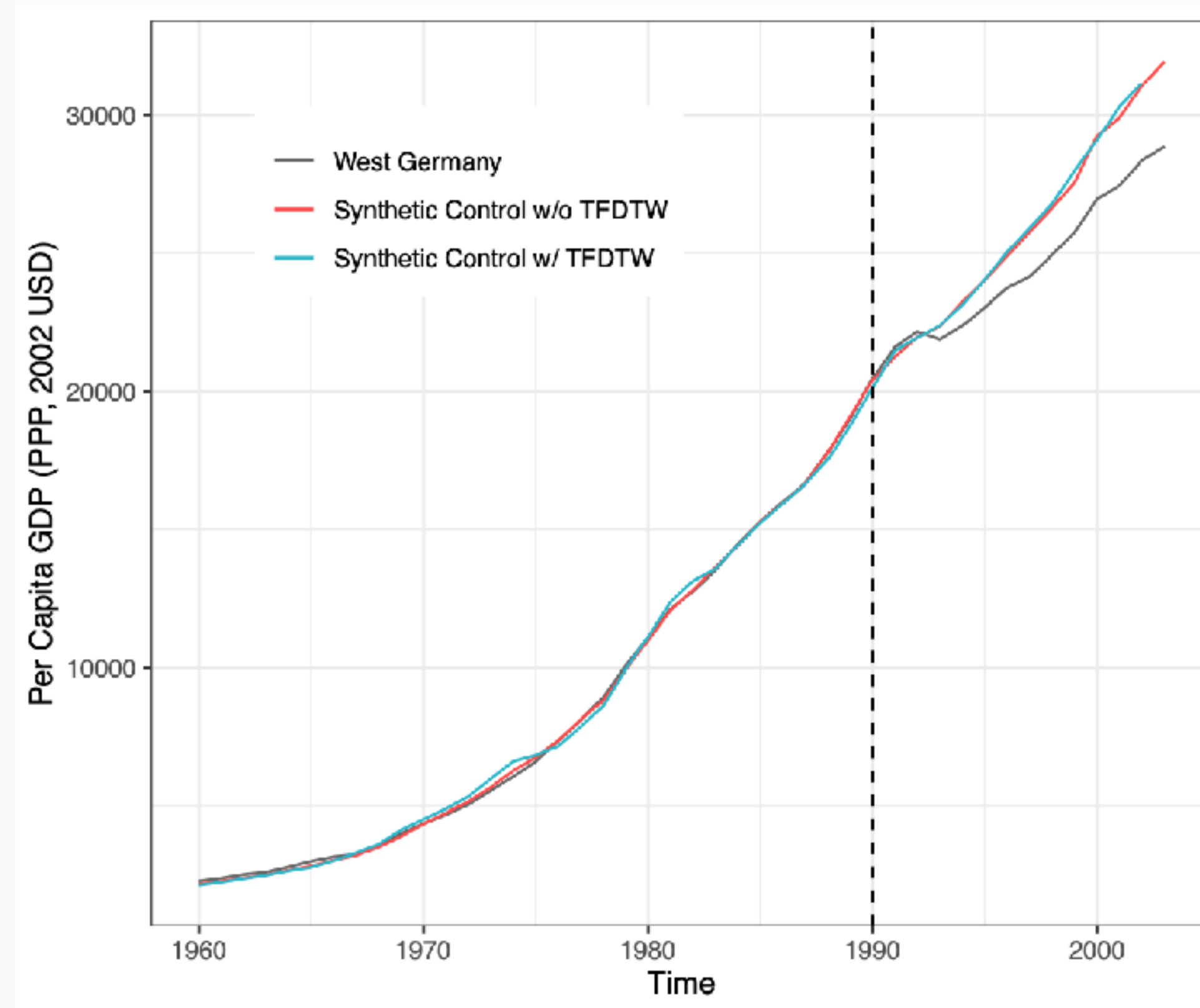
# Results

$$\text{Log Ratio} = \log(MSE_{w/TFDTW} / MSE_{w/oTFDTW})$$



# Does TFDTW Change Abadie's Results?

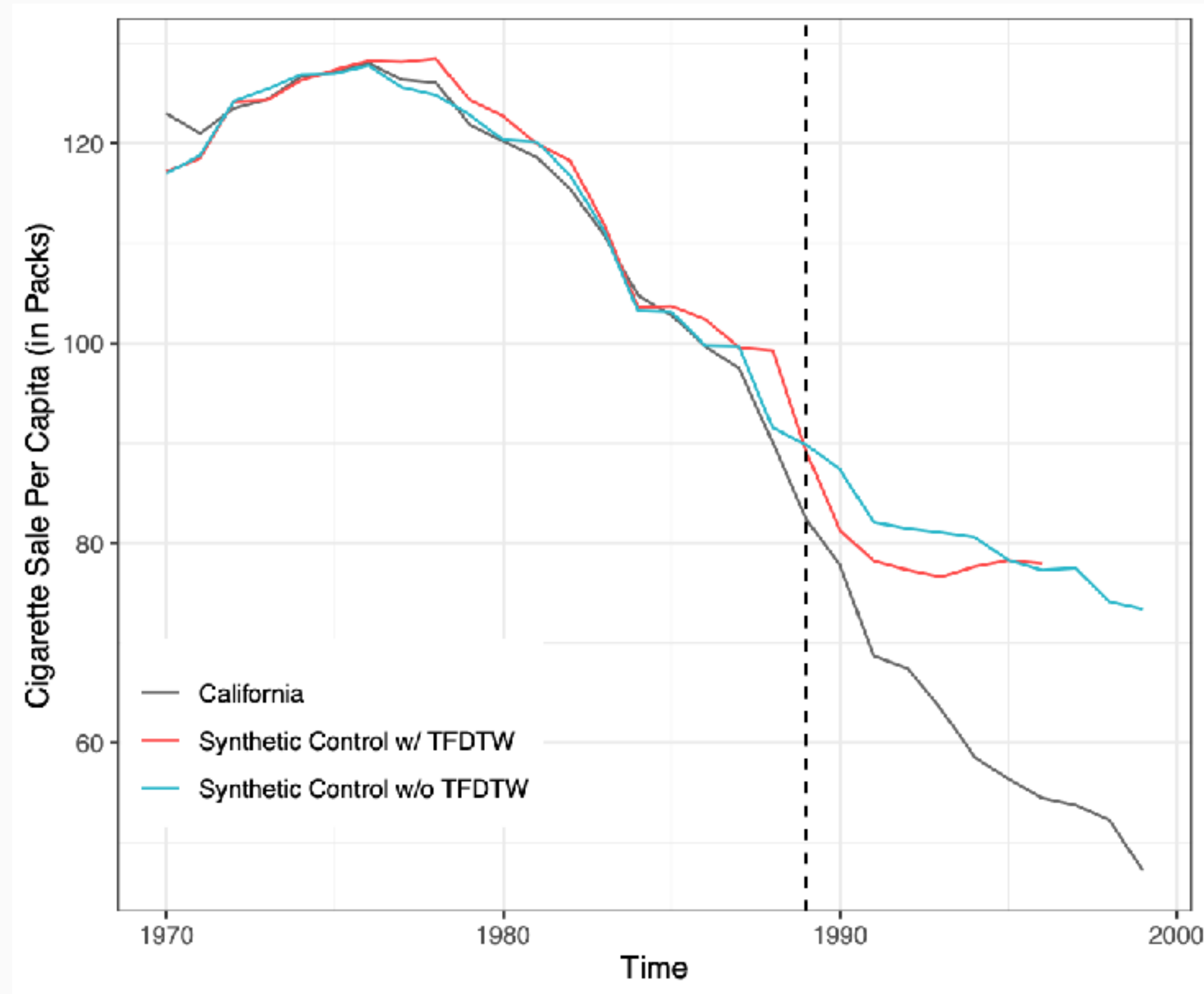
## German Reunification - 1990





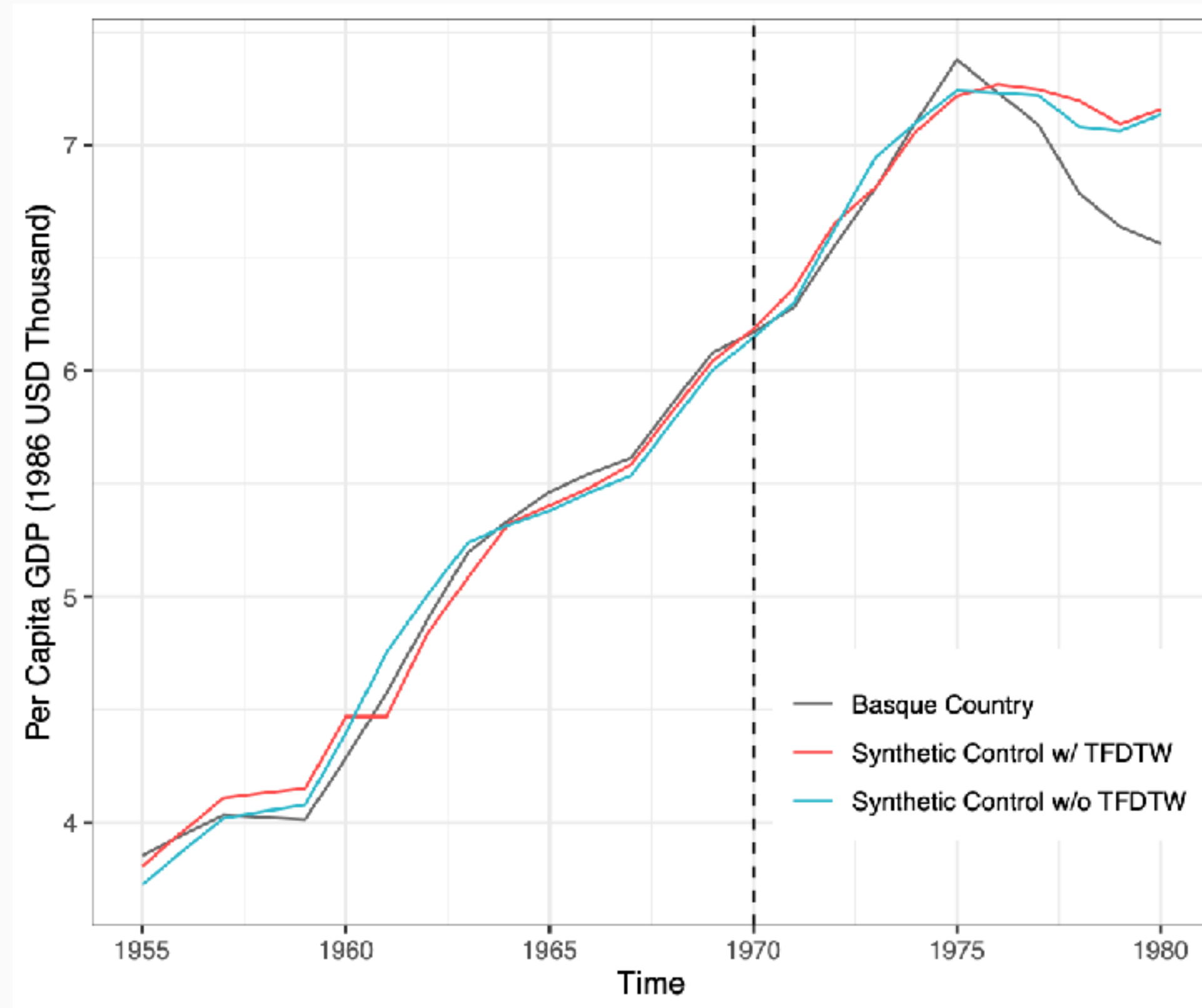
# Does TFDTW Change Abadie's Results?

## California Tobacco - 1989



# Does TFDTW Change Abadie's Results?

## Basque Country, Spain - 1970



# Conclusion

- TFDTW significantly improves synthetic control's ability to account for different speeds in time series.

# Conclusion

- TFDTW significantly improves synthetic control's ability to account for different speeds in time series.
- After applying TFDTW in Abadie's analyses, the estimated causal effects are changed.
  - German reunification ↓
  - California tobacco law ↓
  - Basque country terrorism ↑