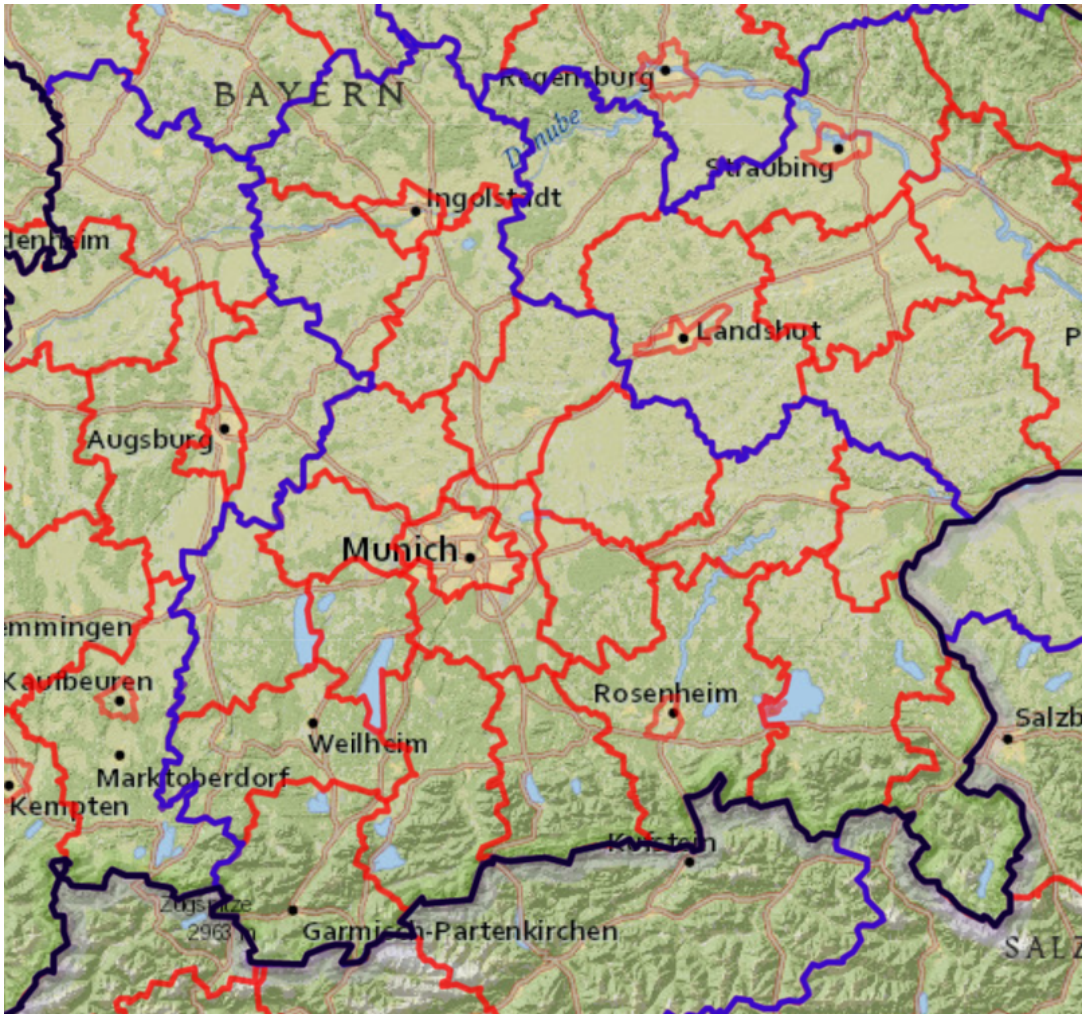
Present situation of the MSc final project

What I planned to do, current situation and plans for project submission

What I planned to do:

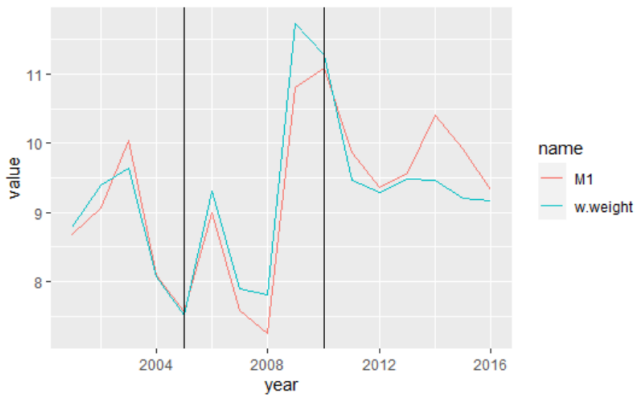
- Have local GDP (outcomes) and average pollution (mediator) in each German city that applied a Low Emission Zone city (represented by a NUTS 3 region). Use a large set of European controls to create a synthetic control for each city and estimate the the treatment effects and ATE. Complement this with the effect labor outcomes for NUTS 2 regions.
- Local GDP was highly affected in the financial crisis and had an important confounder:
    - the German program of vehicle substitution who was the biggest in the word (5billion€, **61€** - 0.15% of GDP - per capita).
- So I focused in Labor outcomes (with lower spatial definition). This greatly limited my number of treated units as many NUTS 2 regions include more than one treated or are too big to really consider them as a good representation of the city's outcomes. From 58 LEZ I end up having 7 large cities that I could study, with only 5 that have a strong dominance of their region.

Münich NUTS 3 (red) and NUTS 2 (blue)



- I planned to use MASC, a method being developed by Giovanni Mellace, Alessandra *Pasquini* to estimate the mediating effect of the reduction of pollution caused by LEZ in labor outcomes. Alessandra nicely provided me with the R code and I had to modify it for my purposes.
- **Pollution estimates** (derived from satellite data and weighted by population and GDP rasters) **were too noisy to perform MASC.** Without mediators, there is no added value on performing MASC over other Synthetic Control methods. I decided to leave MASC for this project and the idea of having an estimation of the mediator effect.

  Example of values for pollution for a Treated city (M1) and it's synthetic control (w.weight). The high variability does not allow to create a meaningful synthetic control. Announcement in 2005, implementation in 2010.



*Note: A separate section at the end shows some results in MASC. I continue to think the method and research question are relevant but I don't feel I have the **time** to get data from a different source.*

- With that in mind, **there was a clear change in questions and methodology:**
  - Questions:
    - Does LEZ affect labor outcomes?
      - Past research has suggested woman labor outcomes are more affected by pollution given they usually take care of vulnerable population (old and young) ¿Can I find some evidence of that?
      - Are sectors that have actively criticized LEZ (such as retail) especially affected by this policy?.
  - Methodology:
    - I started to use the Generalized Synthetic Control (Gsynth) to estimate the effect of LEZ in labor outcomes given it allows for additional controls that the classical SC lacks such as:
      - Unit and time fixed effects,
      - Confidence intervals of treatment effects calculated by a bootstrap procedure.
      - Joint calculation (all treated at once) of the influence of unobserved factors in the outcome, and a Cross-validation procedure to choose how many to include.
      - Weights of control regions can be negative and don't have to add to 1 (less interpretability but more flexibility)
- I am now looking at the effect of LEZ in various outcomes (and there are some **preliminary** results)
  - Unemployment (total and women's unemployment)
    - No changes after announcement, sharp declines after implementation (3-5% less), smaller and sooner effect for female unemployment (1-3% less) (contrary to intuition)
  - $\frac{\text{Woman in active population}}{\text{Men in active population}}$
    - Looks like there is no changes after announcement and then a decline of 3% after implementation, (contrary to intuition)
  - Wages/worker
    - No treatment effect in any city.
  - Proportion of employed population in each sector (especially tourism/retail and public administration)
    - No overall effect in retail overall (looks negative in touristic cities and positive for Hannover - a central German city)
    - No overall effect in public sector share of employment

- **Doubts regarding estimation:** There are 3 main limitations to get the causal effect of the policy on the treated (from less to more difficult to control). All of them are documented by previous research.
  - **(1) Anticipation to the policy** - which is fixed using the announcement date.

- **(2) Partial spillovers of the policy** (cars go in and out of each city and the policy affects surrounding areas too) - Which I control by excluding NUTS regions that have any space in a 30 and 60km radius of a treated city. This highly reduces the number of controls in Germany, Netherlands and Belgium.
- **(3) Time-variant confounders that act after the announcement of the LEZ** (2006, 2007, 2008, 2009 for my sample). For example when comparing German cities unemployment with some other European cities, the effect of the financial crisis is clear. - Which I try to control taking only cities from Germany, Netherlands and Belgium.
  - ❗ I still think that, for unemployment, it is **impossible** to control for the "Scrappage" program mentioned before and the results should be presented not as causal effects of LEZ. Other estimates might be credible as treatment effects of LEZ, but are also questionable.

Current situation

- I feel $\frac{\text{Woman in active population}}{\text{Men in active population}}$, $\frac{\text{Wage}}{\text{Worker}}$ and Proportion of employed population in each sector (especially tourism/retail and public administration) are good outcomes but also **sticky** economic variables that don't change much usually and it is hard a LEZ might change them.
- I feel that, to complete my research (**or even as the principal part**), I can estimate effects in local GDP and GDP/sector of LEZs that were that announced **after** September 2009 (when the scrappage program ended). For that I already researched the announcement dates of most of LEZ and 29+ are after September 2009 and looks like a good treated pool.
  - They are usually small cities that do have sufficient similar-sized controls inside and outside Germany and I have 7+ years of pre intervention period to give confidence that the the pre-intervention paths are parallels.
  - Provisional estimates are usually non-significant until the application of the LEZ. After they show an ATE of 5% decrease in GDP/capita

  ❓ ***Do you think this would be credible? It looks robust to me.***
- There are two ways to estimate my results (which usually end up with very similar results)
  1. Restrict controls to NUTS regions that are farther away of **60km** from any LEZ and keep them for most Europe,
  2. Restrict controls by **30km** distance but only keep Germany or Germany + Netherlands, Belgium and Austria)

  The first might increase overfitting while the second might not fully control for spillover effects of big cities.

  ❓ What do you think about them?

Outline of the capstone project

> *This is just a tentative index, the promised Skeleton of the capstone is in a separate PDF.*

1. Introduction
    1. Reasons for LEZ (Pollution -> health -> legislation -> objective is to reduce pollution)
        - Explain prevalence of pollution, sources and relation with road traffic
        - Explain the institutional background of LEZ, why are they applied and how they are applied in Germany
    2. Economic critiques of LEZ (Economy, Investment,)
        - Economic theory of suboptimal allocation and externalities
        - Examples of critiques (studies and opinions from newspapers)
    3. Why and how (literature) a LEZ can improve the economy? (theory of change)
        - (1) Less congestion, less pollution, less health costs creates...
        - Less mortality, more productivity, better cognitive performance, less absenteeism.
            - Less absenteeism especially in women
    4. My question, does this apply to LEZ? Which effect prevails?
        - Explain the major confounding factor: the application of the scrappage program
2. Data Sources
    1. Y, X: Eurostat
    2. Treatment status of every European city: Ministry of environment, historical documents and scrapped webpages.
3. Analysis techniques
    1. Generalized Synthetic Control Method
        1. Characteristics, reasons to choose model and how do I control for confounders
        2. The effect of LEZ in small cities' aggregate output
            - Selection criteria for treated and control
            - Justify why I believe a causal effect can be derived from this estimates (no major shocks + over 7 years of pre-treatment period)
            - GDP per capita
            - GDP per worker
            - GDP per worker on selected sectors:
                - Retail
                - Transport
                - Public
            - Proportion of GDP of selected sectors
        3. The effect of large LEZ on Labor outcomes
            - Selection criteria for treated and control

        - $\frac{\text{Woman in active population}}{\text{Men in active population}}$ (any positive effect for women?)
        - Employment proportion of relevant sectors
            - Retail sector (does it decrease?)
            - Transport sector (does it decrease?)
            - Public sector (does in increase?)
4. Results
    - *Presented in such a way we the general ATE is the main result.*
    - *We then see the specific treated and synthetic control for some regions that are of special importance or act as good examples.*
        - *The pre-intervention placebo*
        - *Treatment effect*
        - *Map of control regions by weight*
    1. The effect of large LEZ on Labor outcomes
    2. The effect of LEZ in small cities' aggregate output
        - Explain that GDP $\neq$ aggregate utility and having less pollution is surely valuable
5. Conclusions
    1. Summarize, conclude, further research (mention MASC and possibility of doing mediation analysis)
6. Annexes:
    1. Data sources
    2. Implementation on LEZ in Germany (with announcement date)
    3. NUTS2 and 3 regions for Germany and Europe

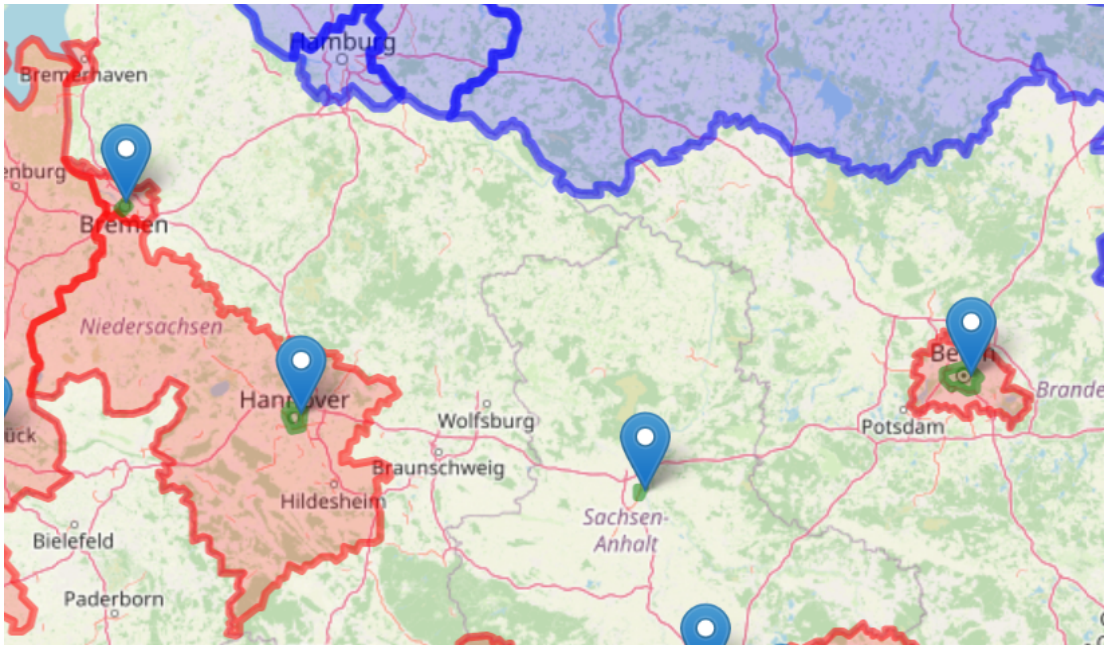
Other topics to discuss:

1. Stay at the Spanish central bank with talks with Laura Hospido about this project.
2. The extension period was granted for 10 working days

End

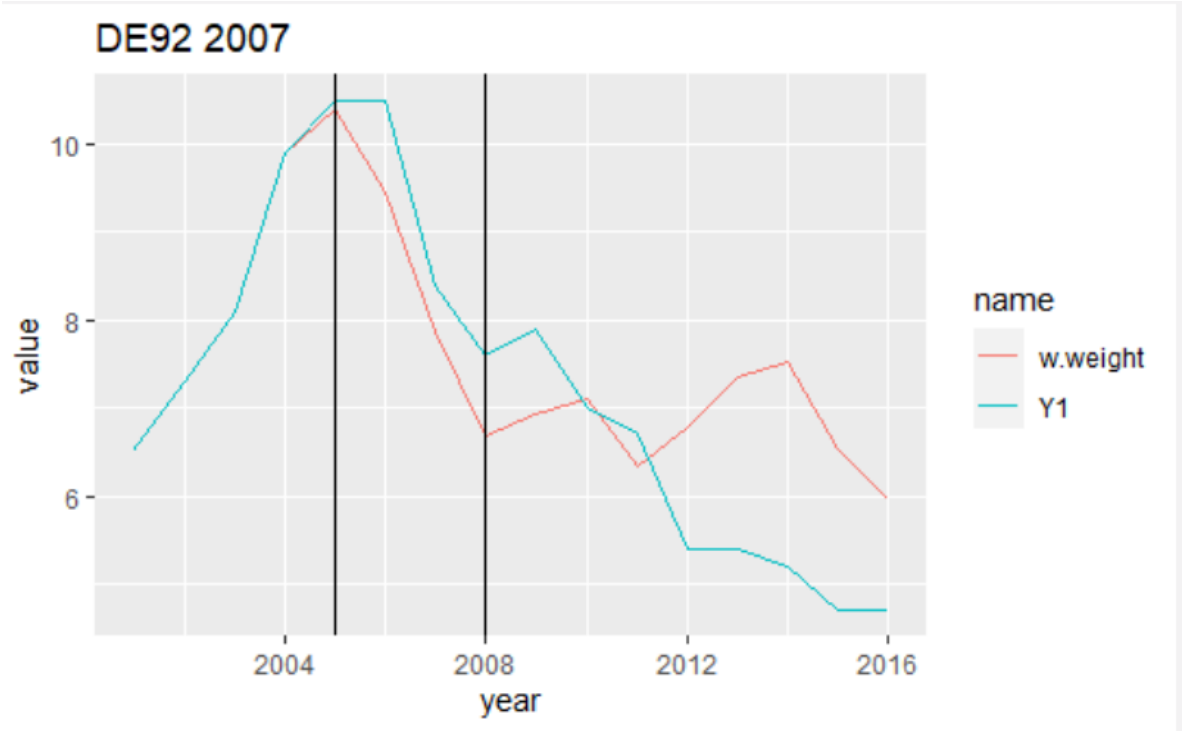
Extra: preliminary results from MASC - If you are interested but not that relevant to my next steps in the capstone :)

Notes on preliminary results - The example of Hannover



Hannover, Bremen and Berlin, It's LEZ and NUTS2 regions

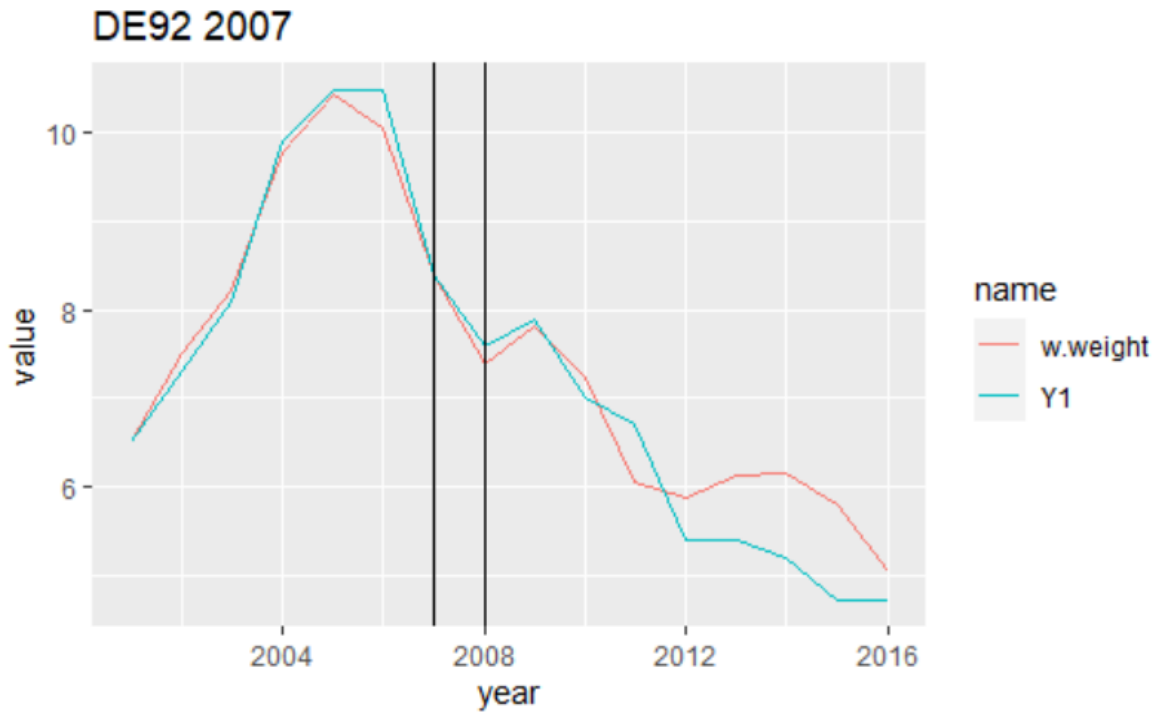
1. Test if I can indeed create a synthetic control with a pre-intervention placebo (the announcement date is 2007 but I act as if it were 2005):



Y1 is the real "Hannover", **w.weight** is the synthetic control. To test that there should be no major discrepancies in the pre intervention period (from 2005 to 2007 given the policy has not passed).

I can confidently do this ***only*** for Hannover for overall unemployment.

2. I then create the "real" synthetic control (with all pre intervention periods until 2007, the year of announcement)

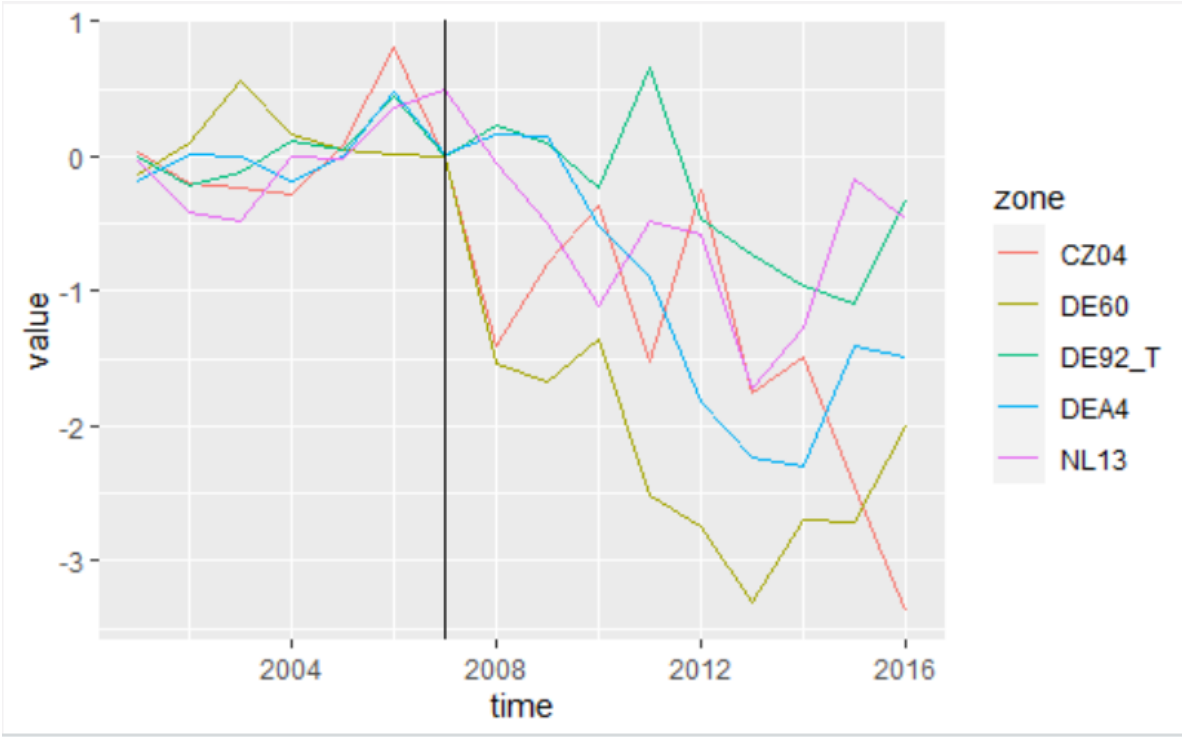


Looks like a good pre-intervention fit. And looks that the LEZ could have reduced unemployment in Hannover.

Furthermore, the control region has very similar pollution levels before the intervention and similar *GDP/capita* and *yearly wages/worker*

	Treated <dbl>	Synthetic <dbl>	Sample Mean <dbl>
special.unempR_T.2001	6.500	6.505	8.547
special.unempR_T.2002	7.300	7.510	8.926
special.unempR_T.2003	8.100	8.224	8.919
special.unempR_T.2004	9.900	9.789	9.240
special.unempR_T.2005	10.500	10.446	9.140
special.unempR_T.2006	10.500	10.060	8.294
special.unempR_T.2007	8.400	8.399	6.939
special.mean_pol_gdp.2001.2007	8.781	8.780	9.590
special.gdp_eur_hab.2007	29700.000	29677.600	21612.245
special.wages_per_worker.2007	32889.369	32943.054	26368.361

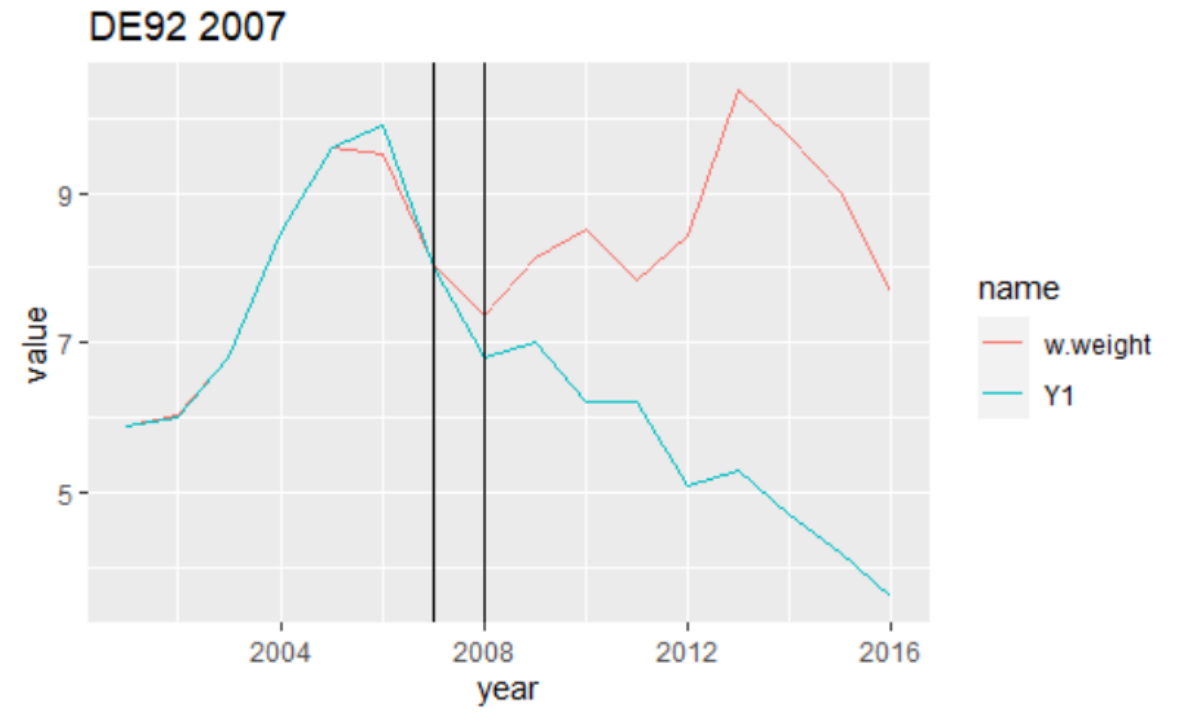
3. Finally I test if this difference is significant by comparing it's magnitude with control regions, especially those that contributed to create the synthetic Hannover. (and restricting to those that have good pre-intervention fit)

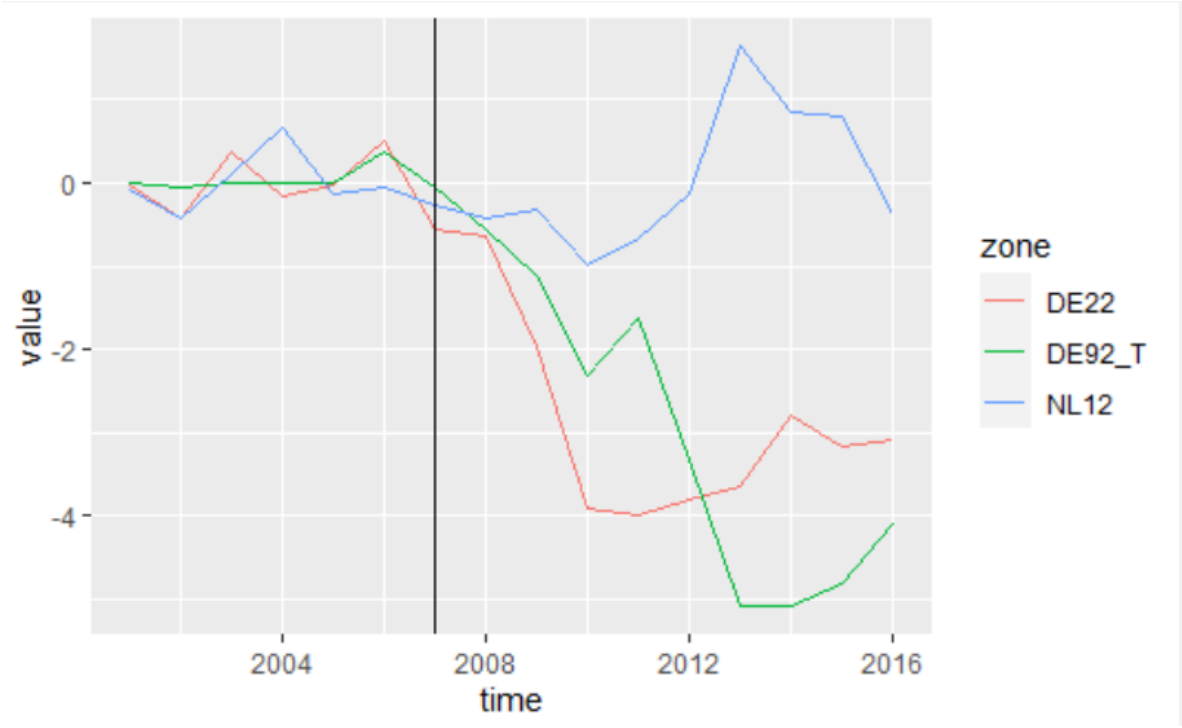


These are the **gaps** between synthetic control and treated (Hannover-DE92_T). result is that Hannover is not that special, All similar control regions experience a reduction in pollution compared to their own synthetic controls.

The effect is not significant.

Female unemployment is similar but more extreme, looks like there was a strong reduction in women's unemployment in Hannover, but it is not specific of that city as other German controls also had the same phenomena:

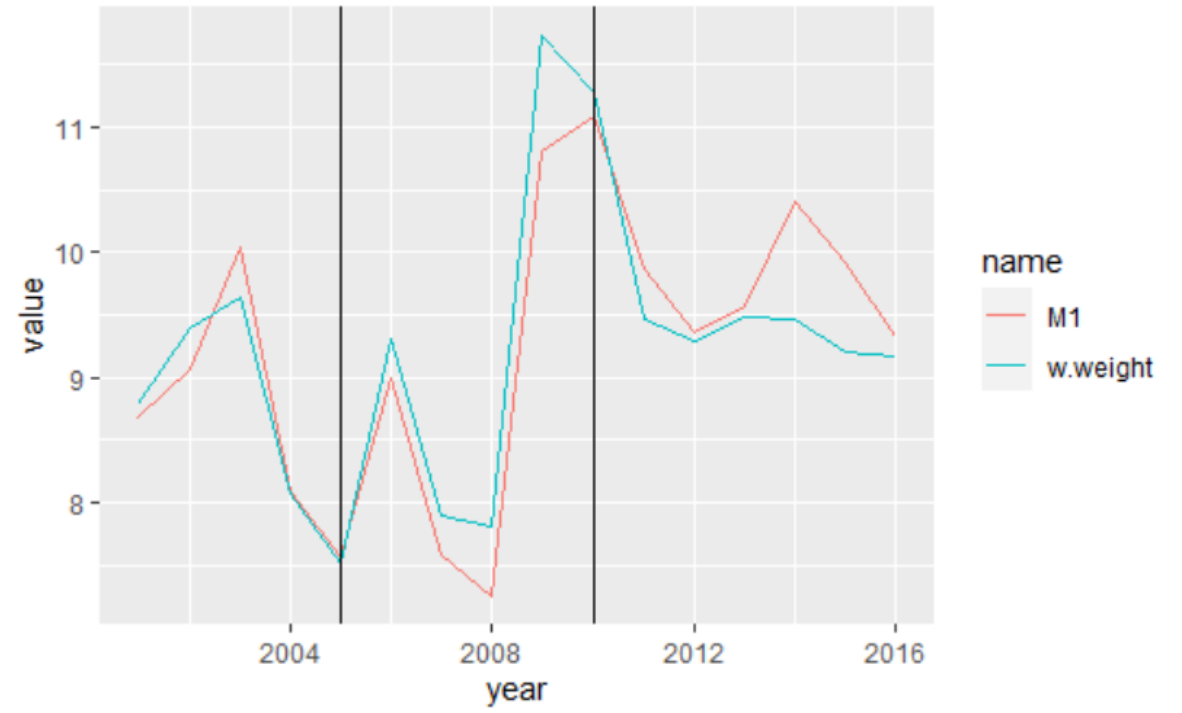




And the Mediator?

On the mediator the results are less interesting, the Synthetic control method fails to find the causal effect of LEZ on pollution reduction even with high quality spatial data. And noise is too big to actually say if there is any effect (or no)

Pre and post-LEZ pollution for Hannover (announcement and implementation marked) - It is like this for **all LEZ**, NUTS 2 and NUTS 3.



Not one result has given interpretable results on the mediator (pollution) after many tries and construction of a pollution measure that specifically targets high productive and high population locations.

As an interesting research question the role of the mediator is a great idea but it does not apply empirically (faulty data or methods - MASC is only a WP which I had to tailor their code to my objectives and could be working imperfectly.)