

MODS206

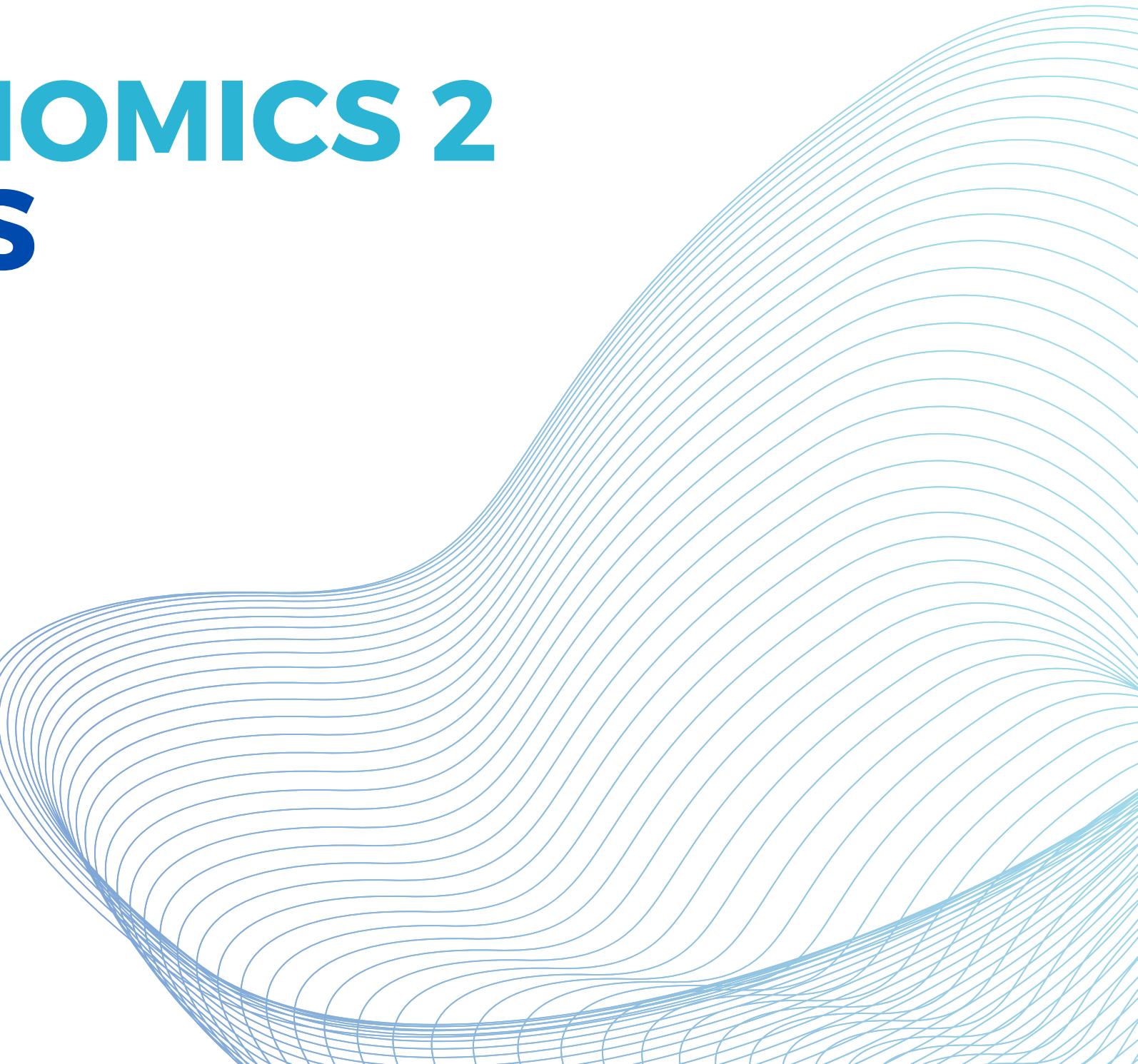
DATA ANALYSIS IN ECONOMICS 2

APPLIED ECONOMETRICS

João Lucas Furtado Melga

Leonardo Hannas C. Santos

Yuri de Sene Alvizi



OUTLINE

- 01 Description of the datasets**
- 02 Exploratory data analysis**
- 03 Causality analysis (diff-in-diff)**
- 04 Conclusion**

DESCRIPTION OF THE DATASETS

DATASET "CITY_INFO"

	city	size	population	popDensity	cityName
0	1	192	379909	1978.692708	Azzurropoli
1	2	91	453991	4988.912088	Biancavilla
2	3	363	271008	746.578512	Lavandonia
3	4	454	152110	335.044053	Celestopoli
4	5	394	52620	133.553299	Plumbeopoli
5	6	183	303659	1659.338798	Aranciopoli

Table 1: Data from each of the six cities

01 Description of the datasets

DATASET "TRAFFIC"

	city	station	weather	dow	datetime	hour	speed	vehicleType	energyConsumption	trafficCongestion	carCrash	pedestrianDeath
0	1	1	3	4	2021-04-01	0	31	2	9.35676	1	False	False
1	1	1	3	4	2021-04-01	0	38	2	8.35296	1	False	False
2	1	1	3	4	2021-04-01	0	29	3	11.29870	1	False	False
3	1	1	3	4	2021-04-01	0	42	1	6.56640	1	False	False
4	1	1	3	4	2021-04-01	0	41	2	7.99116	1	False	False
...
43952647	6	10	1	6	2021-07-31	0	40	1	6.75600	3	False	False
43952648	6	10	1	6	2021-07-31	0	48	2	7.30656	3	False	False
43952649	6	10	1	6	2021-07-31	0	43	2	7.77276	3	False	False
43952650	6	10	1	6	2021-07-31	0	44	3	8.94880	3	False	False
43952651	6	10	1	6	2021-07-31	0	31	1	7.79730	3	False	False

43952652 rows × 12 columns

Table 2: Daily data for each of the ten stations from the six cities

EXPLORATORY DATA ANALYSIS

POPULATION DENSITY OF EACH CITY

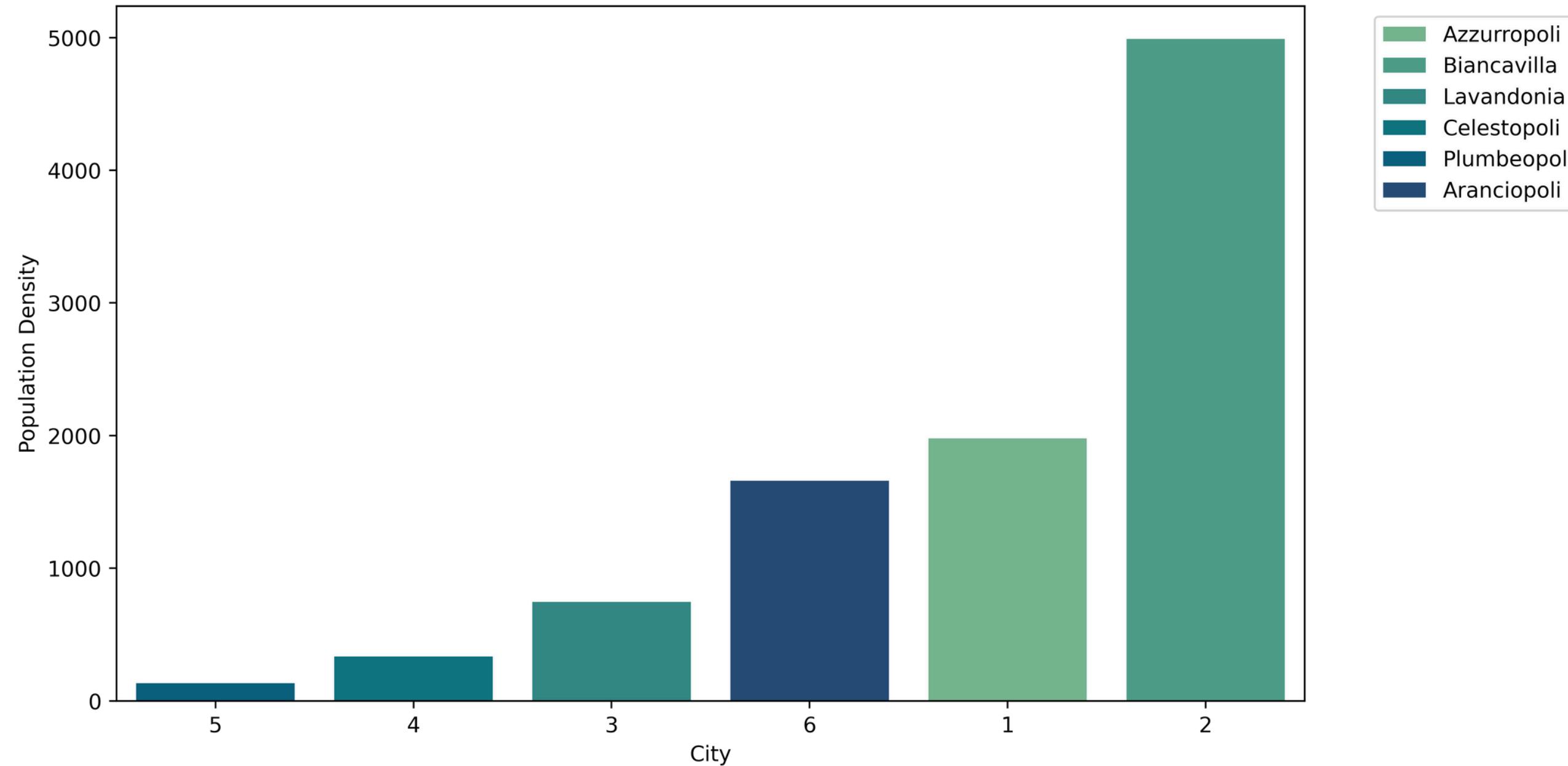


Figure 1: Cities 1 and 6 have similar population densities

MEAN & STD DEVIATION FOR THE CITY 1

city	speed	energyConsumption
1	33.383528	9.552725
2	39.501495	8.566721
3	39.501675	8.566848
4	39.501333	8.565293
5	39.510016	8.566635
6	39.497612	8.567330

Table 3: Mean of speed and energy consumption per city

city	speed	energyConsumption
1	8.901254	2.098720
2	5.007999	1.506273
3	5.008598	1.506875
4	5.009140	1.505781
5	5.007146	1.506842
6	5.008756	1.506719

Table 4: Standard deviation of speed and energy consumption per city

Metrics for the city 1:

- **Speed:** Lowest mean and highest standard deviation
- **Energy consumption:** Highest mean and standard deviation

TEMPORAL DATA VISUALIZATION FOR THE CITY 1

station	datetime	speed	energyConsumption
1	2021-04-01	39.421352	8.586321
	2021-04-02	39.434616	8.586430
	2021-04-03	39.523284	8.562825
	2021-04-04	39.538990	8.565154
	2021-04-05	39.525909	8.564945
...
10	2021-07-27	24.478435	10.987449
	2021-07-28	24.520537	11.008755
	2021-07-29	24.420923	10.987493
	2021-07-30	24.469350	10.965828
	2021-07-31	24.422000	11.069551

There is a clear change on the daily average speed and energy consumption, according to the station and the time window

Table 5: Daily mean of speed and energy consumption per station in City 1

TEMPORAL DATA VISUALIZATION FOR THE CITY 1

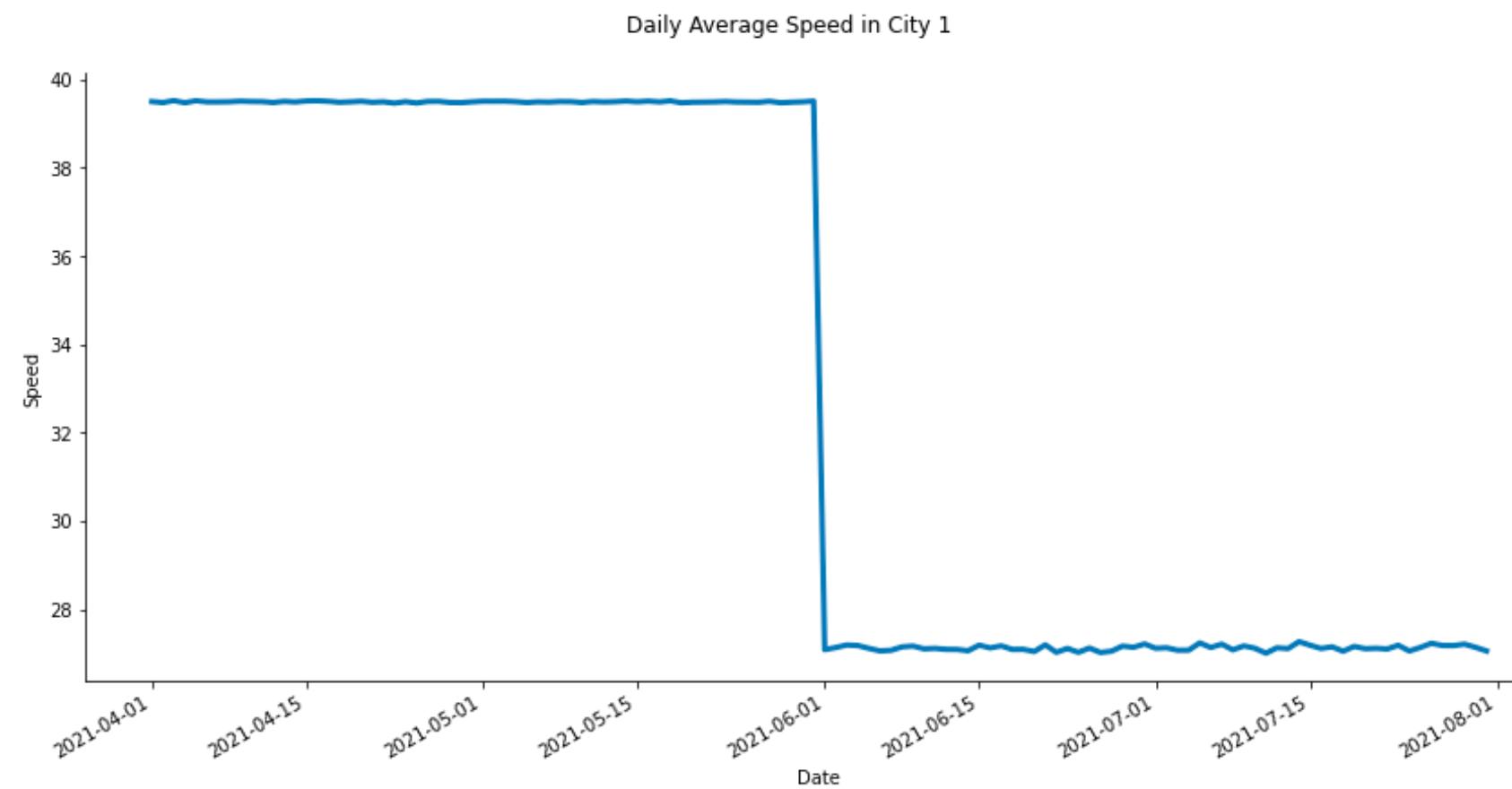


Figure 2: Reduction in the daily average speed in City 1 since June 1st, 2021

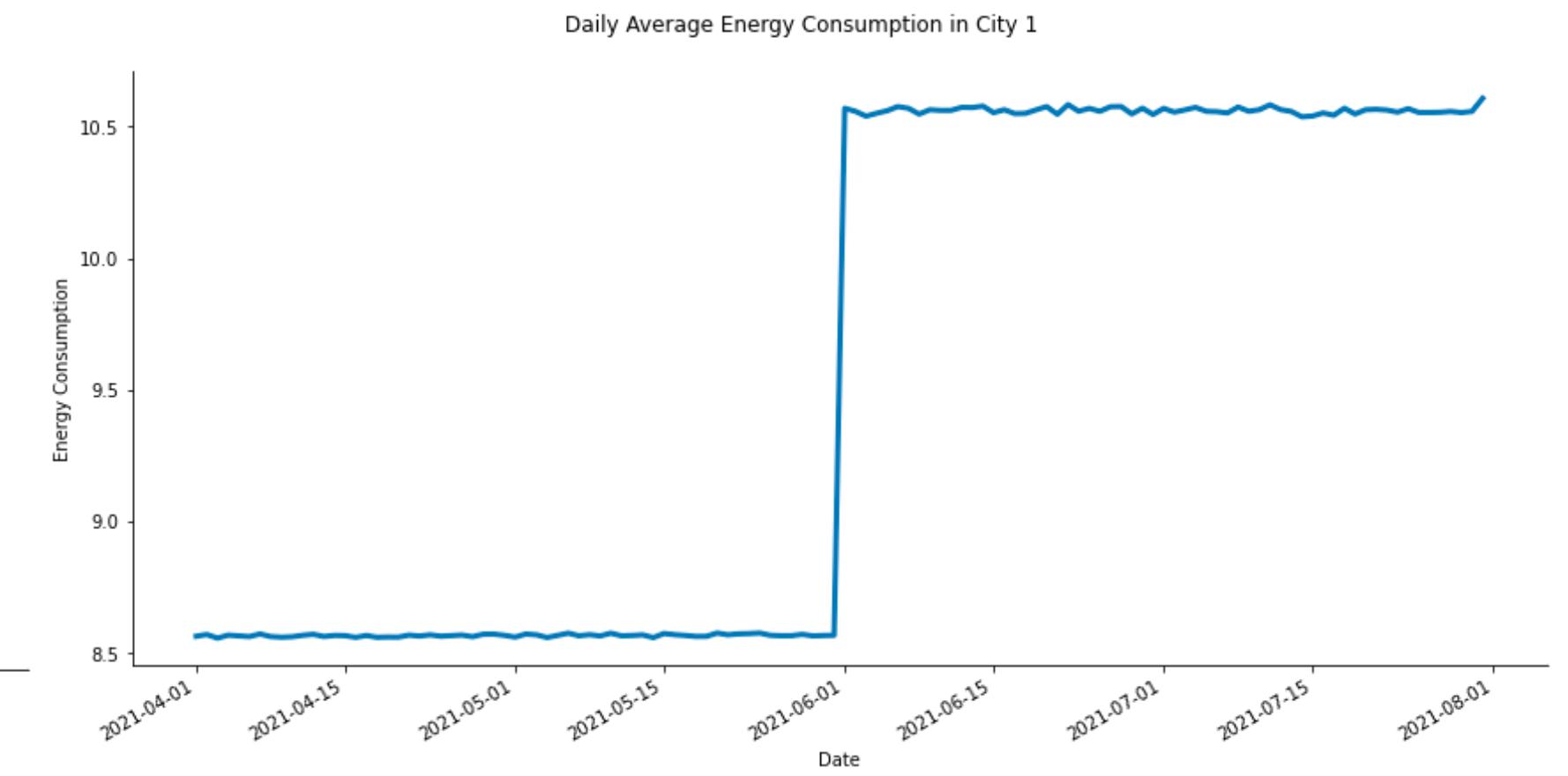


Figure 3: Increase in the daily average energy consumption in City 1 since June 1st, 2021

DAILY AVERAGE SPEED PER STATION IN CITY 1

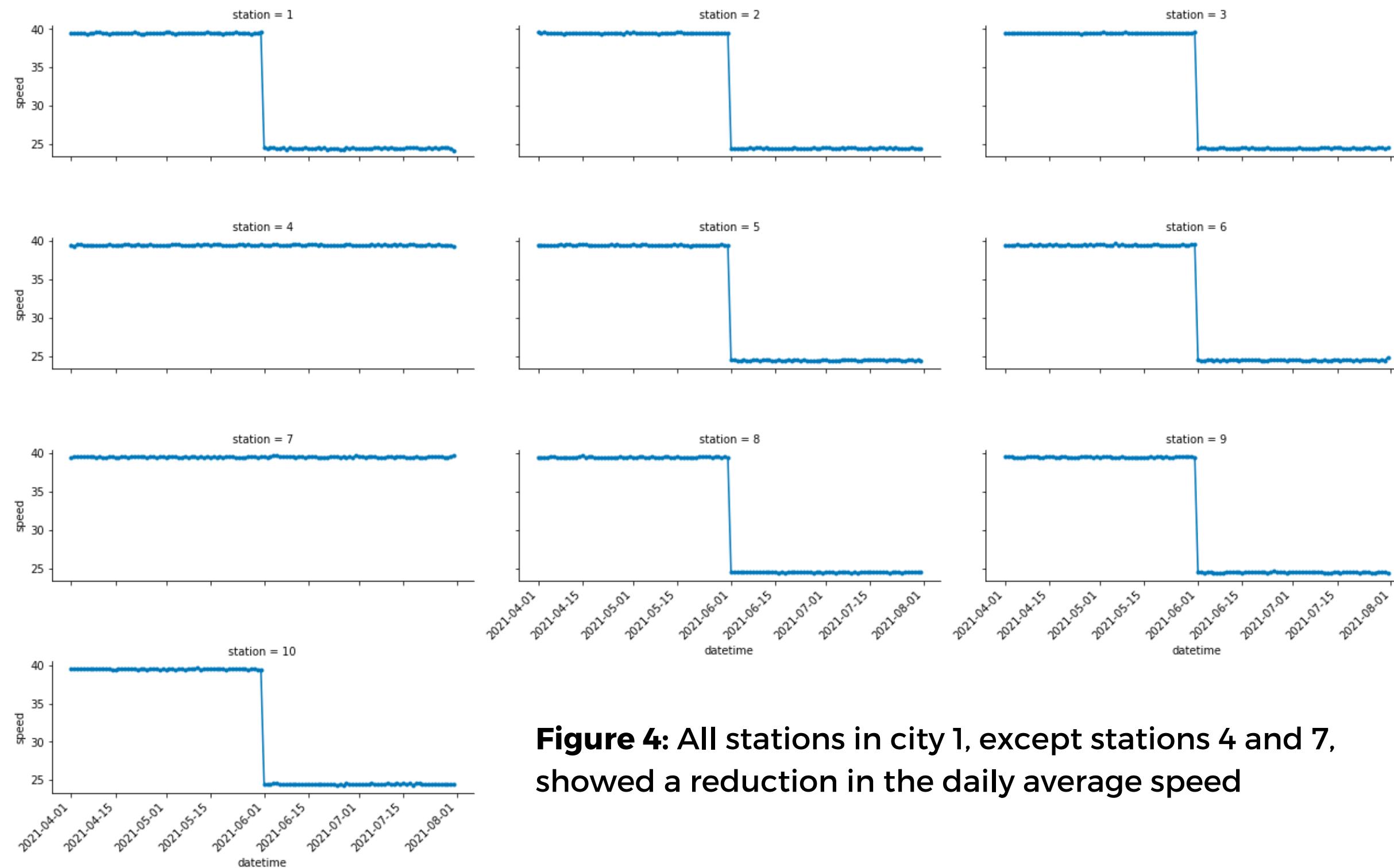


Figure 4: All stations in city 1, except stations 4 and 7, showed a reduction in the daily average speed

COMPARISON: CITY 1 VS THE OTHER CITIES

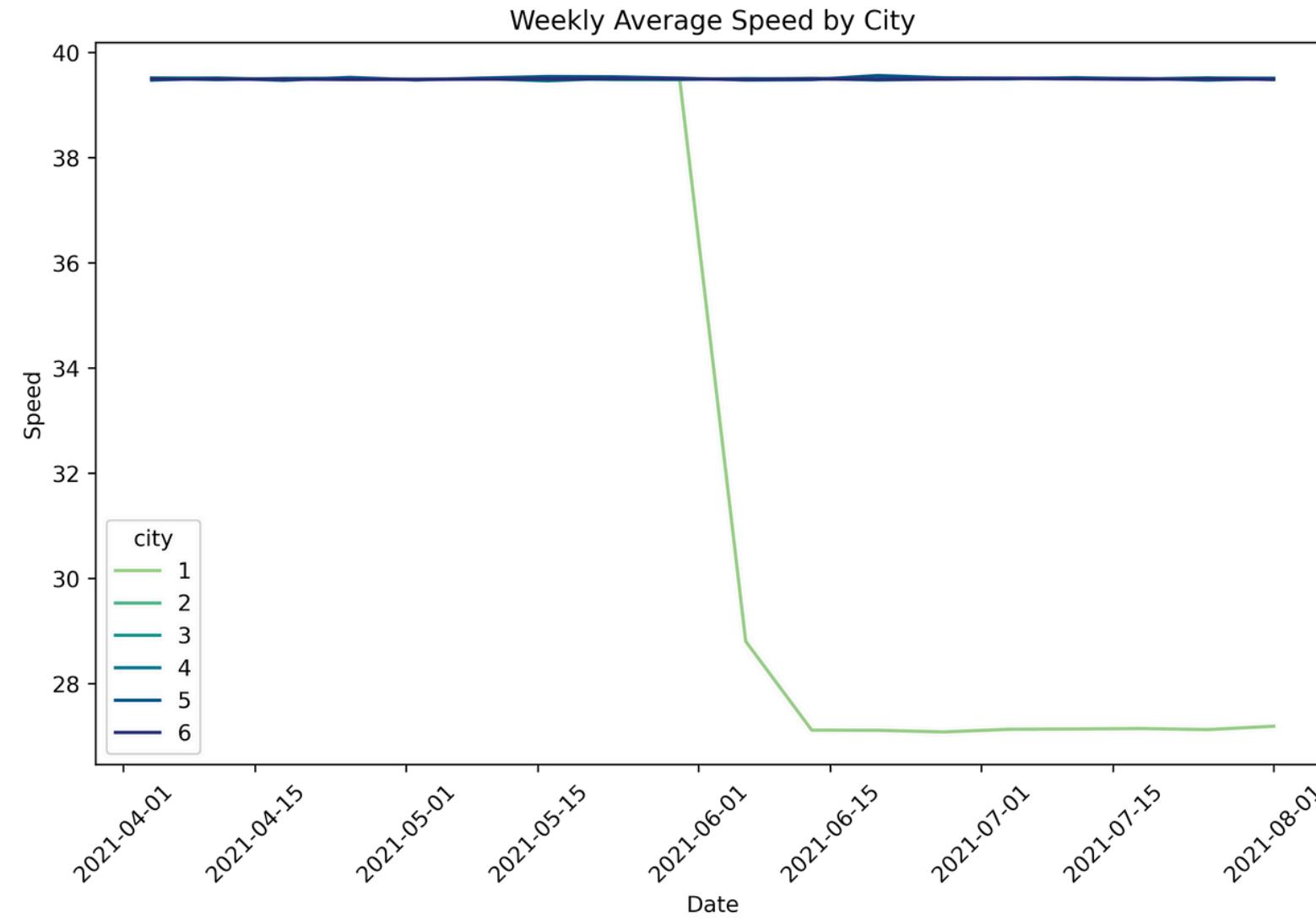


Figure 5: City 1 was the only city that presented a reduction in the weekly average speed among all cities

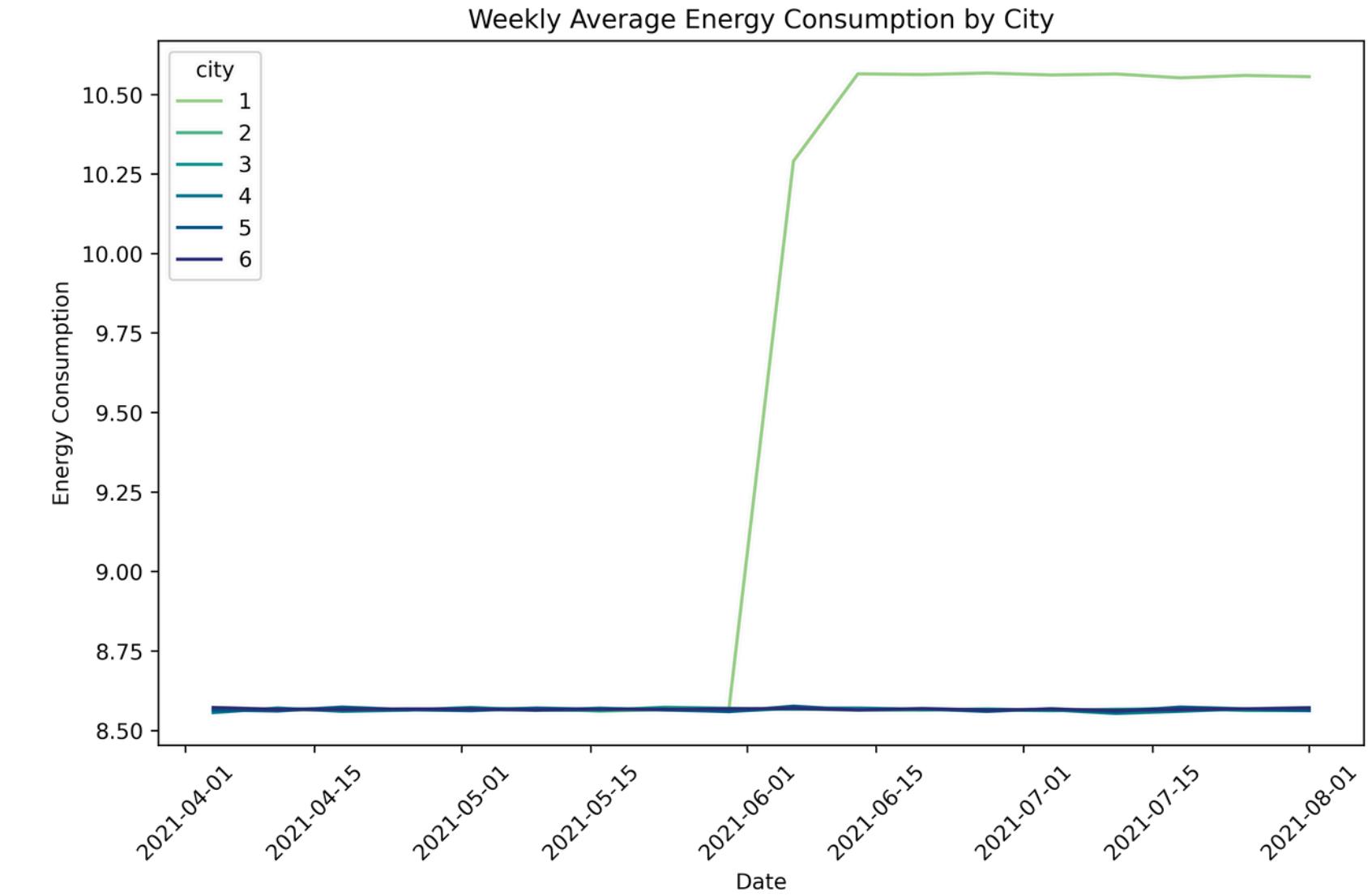


Figure 6: City 1 was the only city that showed an increase in the weekly average energy consumption among all cities

COMPARISON: CITY 1 VS THE OTHER CITIES

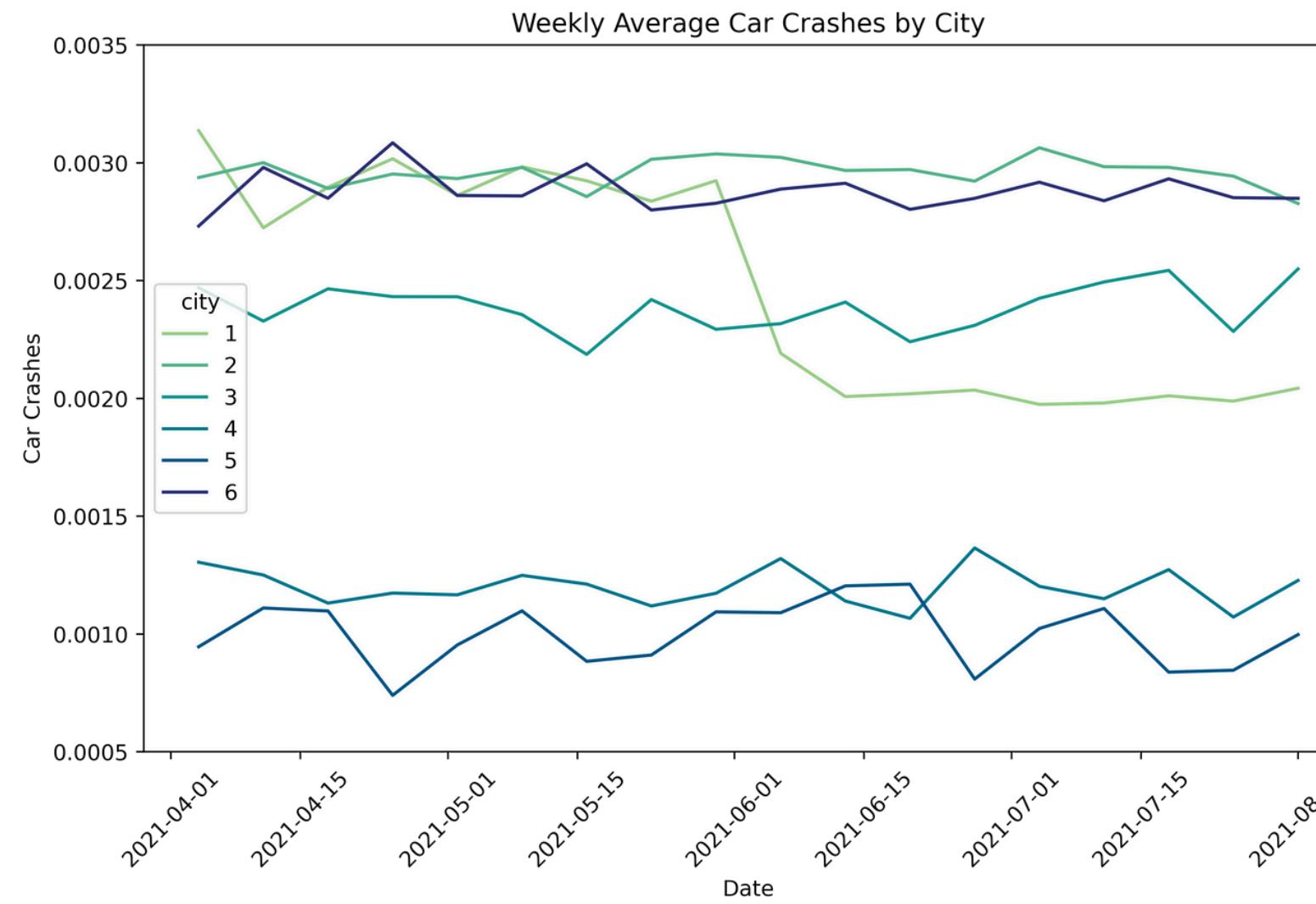


Figure 7: Since the date of change (June 1st, 2021), the weekly average of car crashes has decreased in City 1

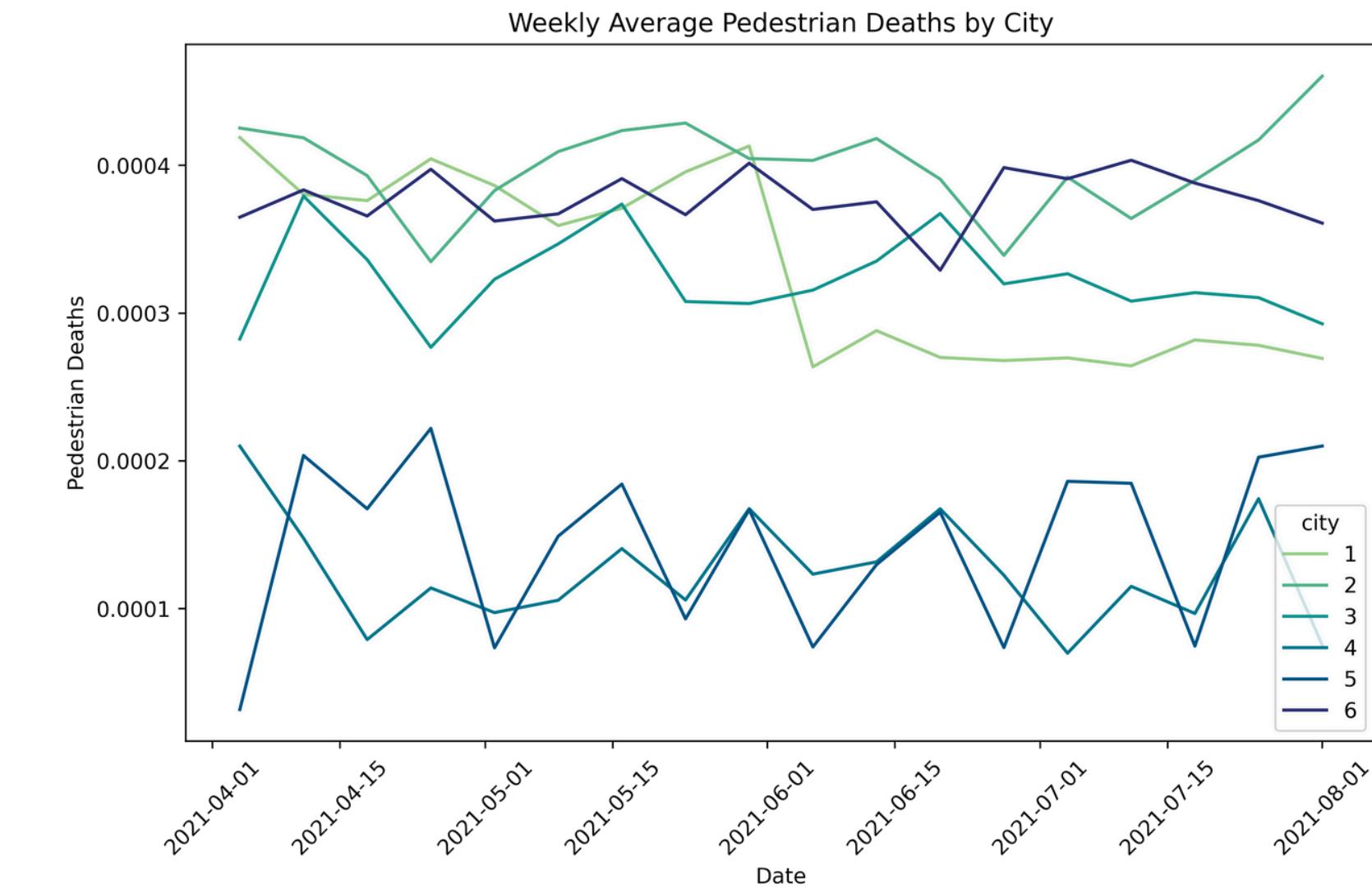


Figure 8: Since the date of change, the weekly average of pedestrian deaths has also decreased in City 1

IN SUMMARY

- Observations:
 - Significant changes in traffic metrics were observed across all City 1 stations (excluding stations 4 and 7) on June 1st, 2021
- Hypothesis:
 - It is hypothesized that City 1 implemented new speed limit regulations in various districts to reduce automobile accidents and pedestrian fatalities
- Further investigation:
 - Additional data analysis is required to ascertain whether the changes in speed limits have effectively reduced car crashes and pedestrian fatalities

CAUSALITY ANALYSIS (DIFF-IN-DIFF)

DIFF-IN-DIFF METHOD

- The Differences-in-Differences (DiD) method is a robust quantitative approach used to estimate the causal effect of a policy intervention or treatment on a targeted metric
- Regression:

$$Y = \beta_0 + \beta_1 \times Treated + \beta_2 \times Post + \beta_3 \times Treated \times Post$$

DIFF-IN-DIFF METHOD

$$Y = \beta_0 + \beta_1 \times Treated + \beta_2 \times Post + \beta_3 \times Treated \times Post$$

- Dependent Variable (Y): Metrics affected by the treatment (e.g., pedestrian deaths, car crashes)
- Post: Dummy variable indicating post-treatment period
- Treated: Dummy variable indicating whether the observation is in treatment or control group
- β_0 (Intercept): Baseline level of the dependent variable
- β_1 (Treatment Effect): Represents the effect of being in the treatment group
- β_2 (Time Effect): Represents the overall time effect across all groups
- β_3 (DiD Estimator): Key coefficient estimating the causal impact of the treatment post-intervention relative to before

DIFF-IN-DIFF METHOD

- Implementation of Tests:
 - Within City Analysis: Compare stations within City 1, using stations 4 and 7 as controls
 - Between Cities Analysis: Compare City 1 (treated) with City 6 (control), taking into account their similar population densities
- Preparation for DiD Analysis:
 - Dummy ‘treated’: Assign ‘0’ for the control group and ‘1’ for the treatment group
 - Dummy ‘post’: Assign ‘0’ for observations before June 1st, 2021 and ‘1’ for observations after this date

DIFF-IN-DIFF REGRESSION

- Within City 1 analysis:

Dep. Variable:	carCrash	log(carCrash)
	coef (std err)	coef (std err)
Weekly average	const	0.002858 (4.66e-05) -5.86 (1.99e-02)
	post	8.219e-05 (6.59e-05) 0.02774 (2.82e-02)
	treated	7.023e-05 (5.21e-05) 0.02466 (2.23e-02)
Post-treatment effect	interaction	-0.001183 (7.37e-05) -0.5021 (3.15e-02) $\Delta Y = \beta_3 * 100 = -50.21\%$
	Observations	180.0 180.0
	R-squared	0.882 0.884
	Adj. R-squared	0.880 0.882

DIFF-IN-DIFF REGRESSION

- Within City 1 analysis:

Dep. Variable:	pedestrianDeath	log(pedestrianDeath)
	coef (std err)	coef (std err)
Weekly average	const	0.0004013 (1.96e-05) -7.846 (5.72e-02)
	post	-1.081e-05 (2.77e-05) -0.039 (8.09e-02)
	treated	-1.592e-05 (2.19e-05) -0.03532 (6.39e-02)
Post-treatment effect	interaction	-0.0001212 (3.10e-05) -0.3976 (9.04e-02) $\Delta Y = \beta_3 * 100 = -39.76\%$
	Observations	180.0 180.0
	R-squared	0.395 0.450
	Adj. R-squared	0.385 0.440

DIFF-IN-DIFF REGRESSION

- Between cities analysis:

Dep. Variable:	carCrash	log(carCrash)
	coef (std err)	coef (std err)
Weekly average	const	0.002877 (3.45e-05) -5.855 (1.38e-02)
	post	2.29e-05 (4.88e-05) 0.007618 (1.95e-02)
	treated	3.683e-05 (4.88e-05) 0.01459 (1.95e-02)
Post-treatment effect	interaction	-0.000887 (6.91e-05) -0.3815 (2.76e-02) $\Delta Y = \beta_3 * 100 = -38.15\%$
	Observations	360.0 360.0
	R-squared	0.559 0.598
	Adj. R-squared	0.556 0.595

DIFF-IN-DIFF REGRESSION

- Between cities analysis:

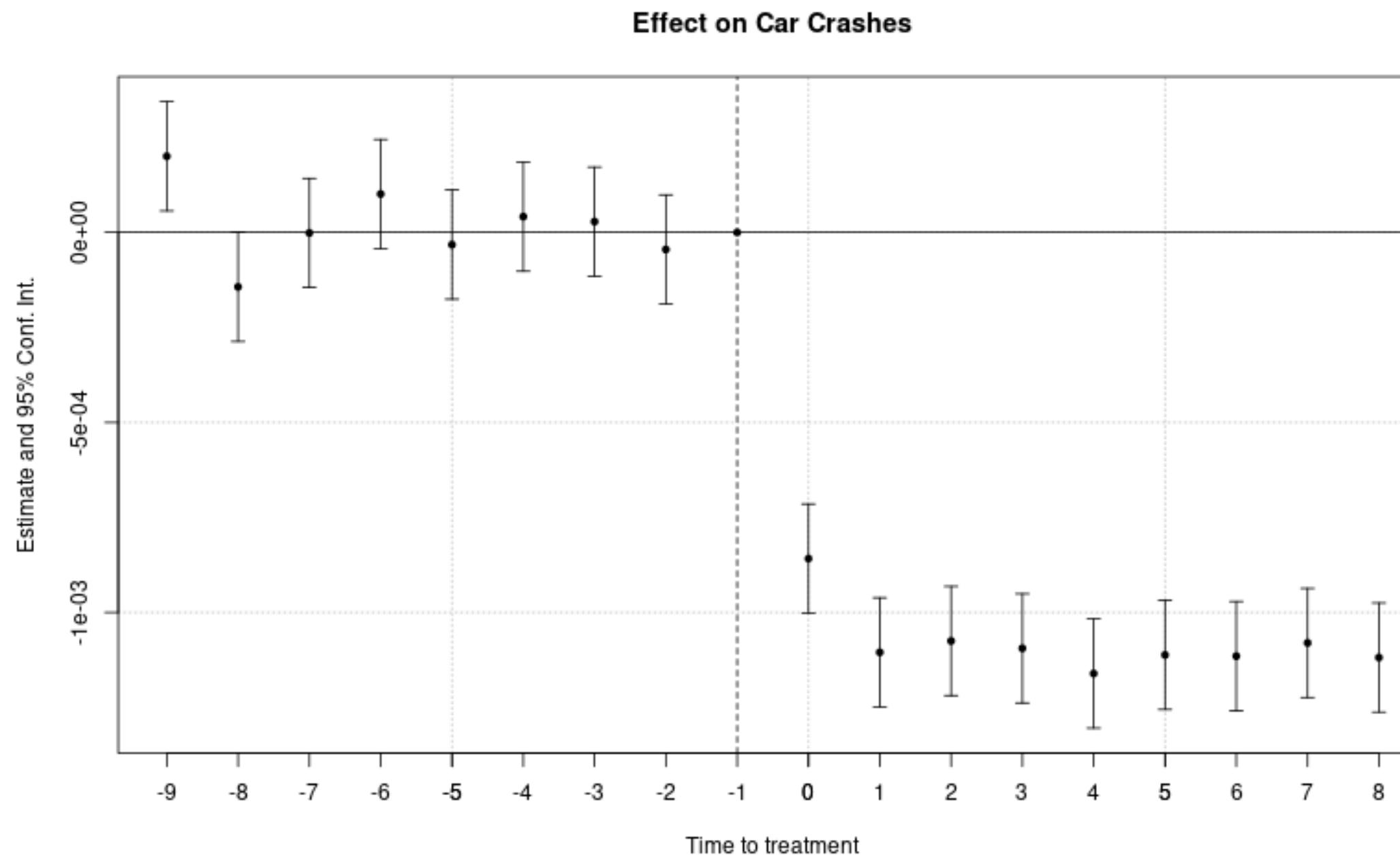
Dep. Variable:	pedestrianDeath	log(pedestrianDeath)
	coef (std err)	coef (std err)
Weekly average	const	0.0003734 (9.45e-06) -7.917 (2.71e-02)
	post	2.678e-08 (1.34e-05) -0.008152 (3.83e-02)
	treated	1.519e-05 (1.34e-05) 0.04208 (3.83e-02)
Post-treatment effect	interaction	-0.0001078 (1.89e-05) -0.349 (5.41e-02) $\Delta Y = \beta_3 * 100 = -34.9\%$
	Observations	360.0 360.0
	R-squared	0.187 0.238
	Adj. R-squared	0.180 0.231

EVENT STUDY

- Approach:
 - To assess the effects of speed limit changes, we conduct a **weekly differences-in-differences regression analysis**
 - Regression used a ‘time to treatment’ dummy variable instead of ‘post’, ranging from -9 to 8 to denote weeks relative to the week of June 1st, 2021.
- Visualization Strategy:
 - For each week, we calculate the β_3 coefficients, representing the estimated impact of the speed limit changes, and plot them along with their confidence intervals

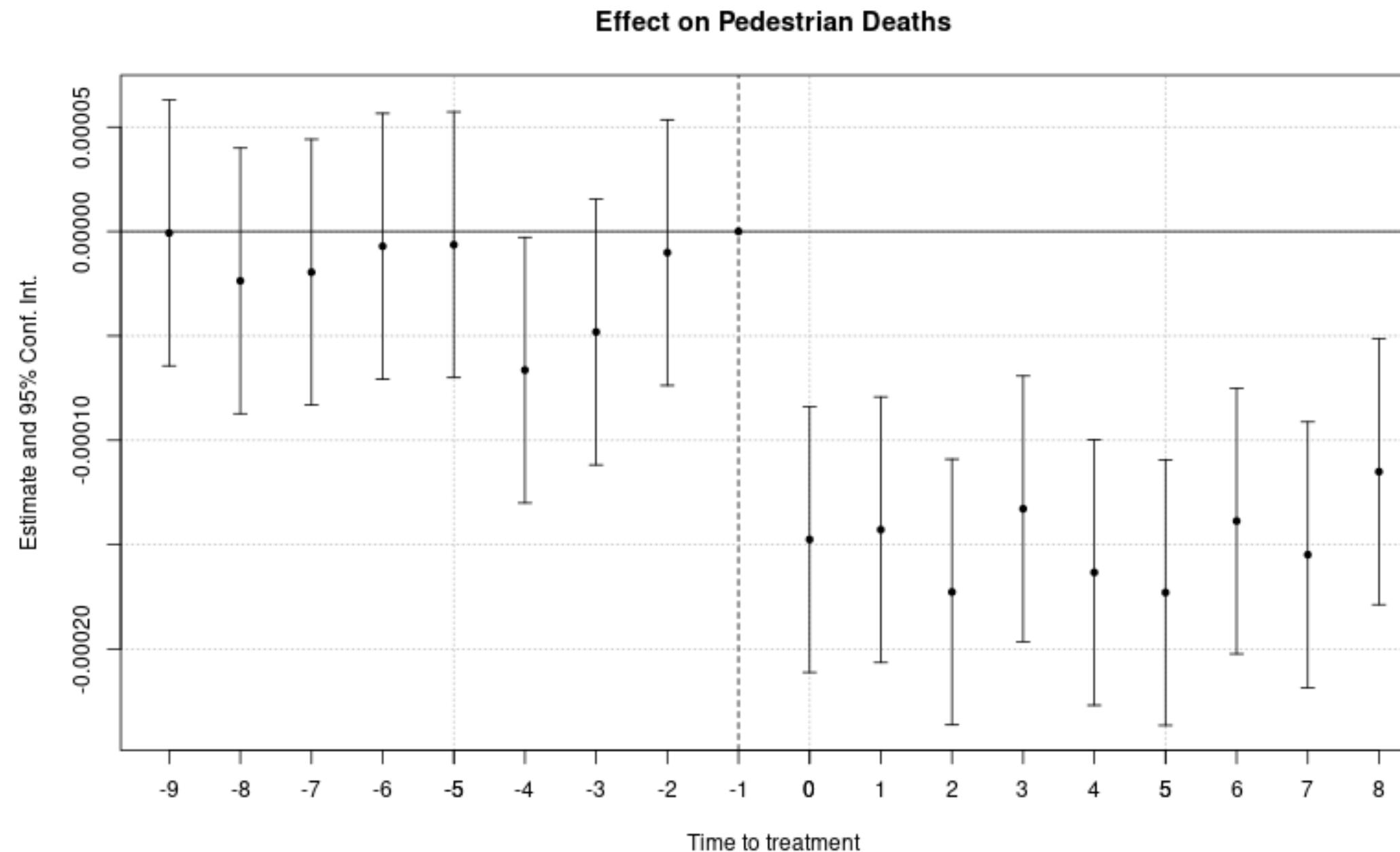
EVENT STUDY RESULTS

- Within City 1 analysis:



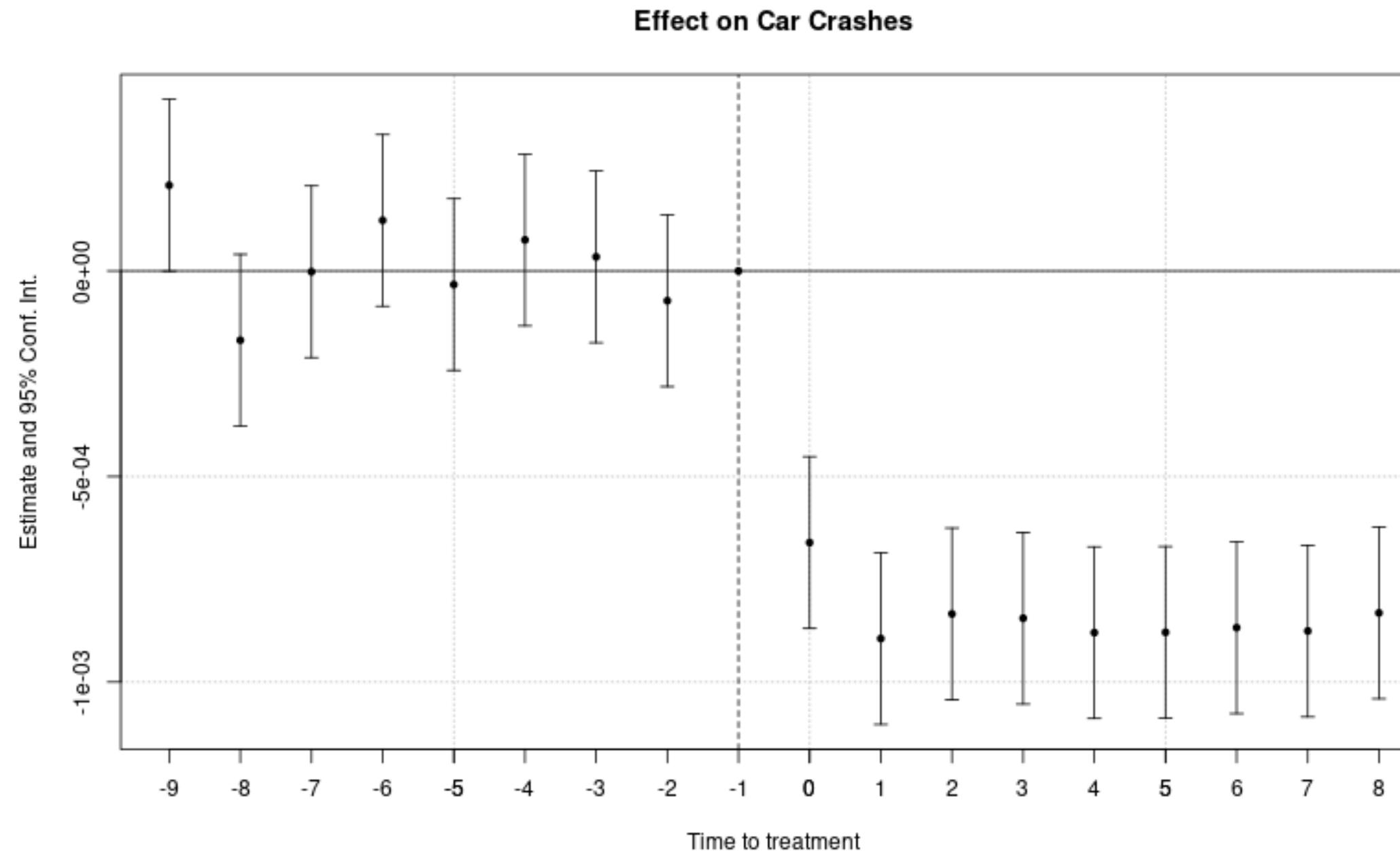
EVENT STUDY RESULTS

- Within City 1 analysis:



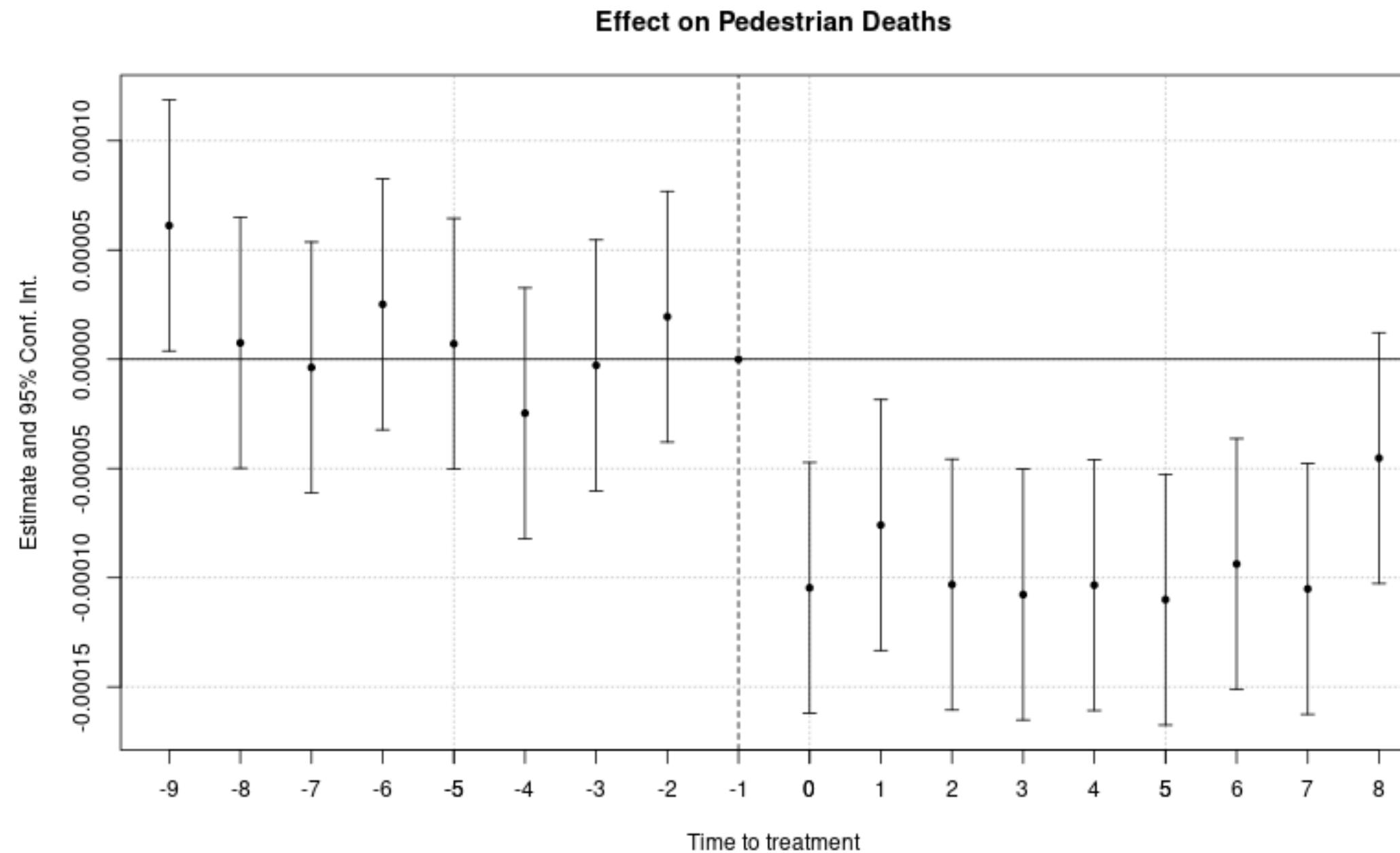
EVENT STUDY RESULTS

- Between cities analysis:



EVENT STUDY RESULTS

- Between cities analysis:



CONCLUSION

MAIN FINDS

- Speed limit has changed on June 21st, 2021, leading to changes in other traffic metrics
- Policy changes in traffic are normally taken to improve safety, what is measured in our dataset in pedestrian deaths and car crashes
- According to the Differences-in-Differences analysis performed, it is possible to affirm that decreasing the speed limit in some stations of City 1 leaded to a significant decreasing in car crashes and pedestrian deaths

THANK YOU!

