# **Staggered Difference-in-Differences in Practice:** Causal Insights from the Music Industry

NAZLI M. ALAGÖZ

# About me

- **Who I Am**
  - 🗣️ Nazli Alagöz (call me Naz!)
  - 🎓 PhD Candidate in Quantitative Marketing
  - 📈 Background: Economics & Econometrics
  - 📅 Graduating: Dec 2023
- **Areas of Interest**
  - 📊 Data Science
    - 👨‍🔬 Causal Inference

# Motivating Example


billboard

CHART BEAT

**Kate Bush's 'Running Up That Hill' Tops Both Billboard Global Charts**

Bush's hit rebounds for a second week at No. 1 on the Global 200 and claims its first frame atop Global Excl. U.S., sparked by its sync in "Stranger Things," whose latest episodes arrived July 1.

By Gary Trust
07/11/2022


Opportunity and uncertainity in the music industry around "music syncronisations" (sync)

# Approach

- What's the impact of music sync on song popularity?
- Challenges in experimentation
  - ❌ Standard A/B testing is impractical in this context
- Instead, use observational data
  - Dataset: Features songs that are both synced and not synced
  - Metrics: Number of playlist placements as proxy for popularity

# How to Estimate the Effect?

- ❌ Try to compare outcomes of synced vs. unsynced songs
  - Ignores the baseline differences in synced vs. unsynced songs
  - Other song-level characteristics
- ❌ Try to compare before and after for the synced songs
  - Ignores time trends and seasonality
- ✅ The solution: Difference-in-Differences (DiD)

# Difference-in-Differences (DiD)

<u>Song</u>

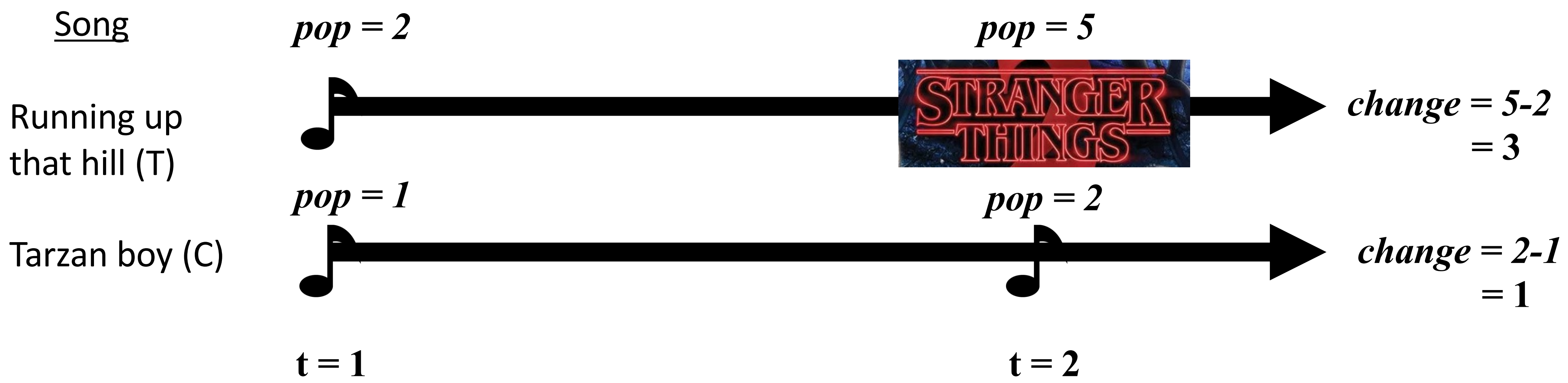
Running up that hill (T)

Tarzan boy (C)


**t = 1** **t = 2**

***Two-group and two-period case***

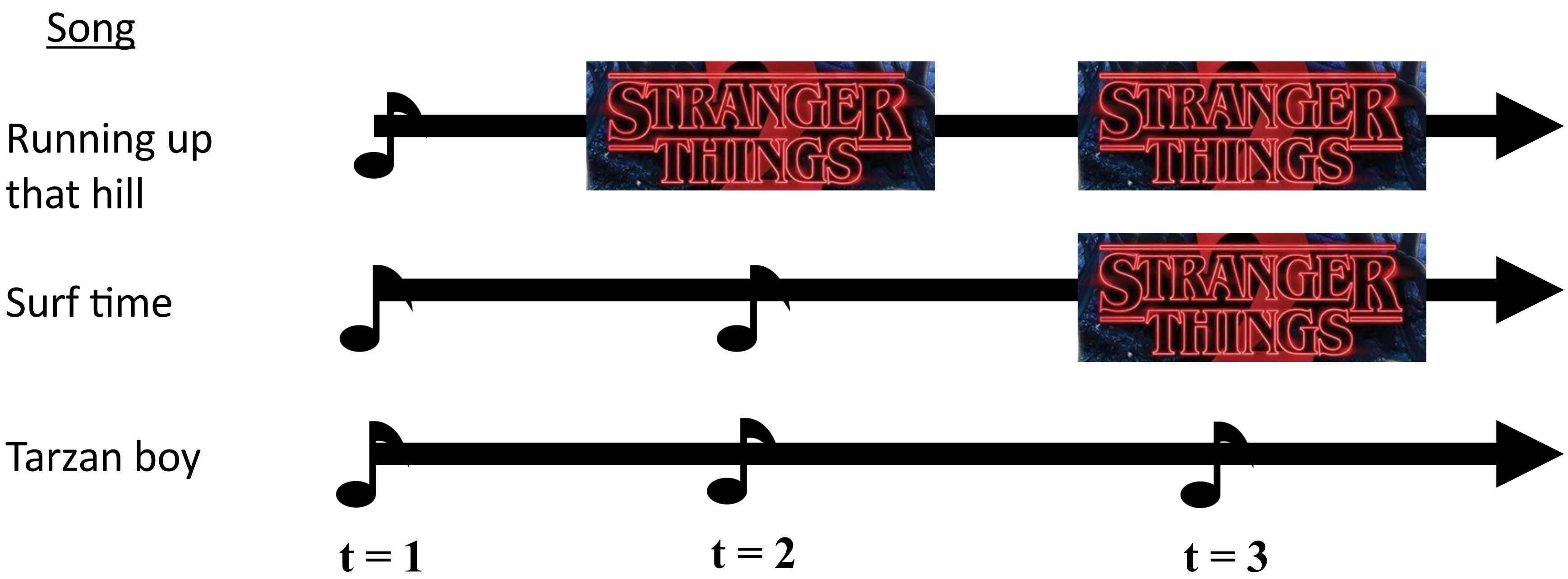
# Difference-in-Differences (DiD)




*Using parallel trends assumption, we can estimate the treatment effect on the treated.*

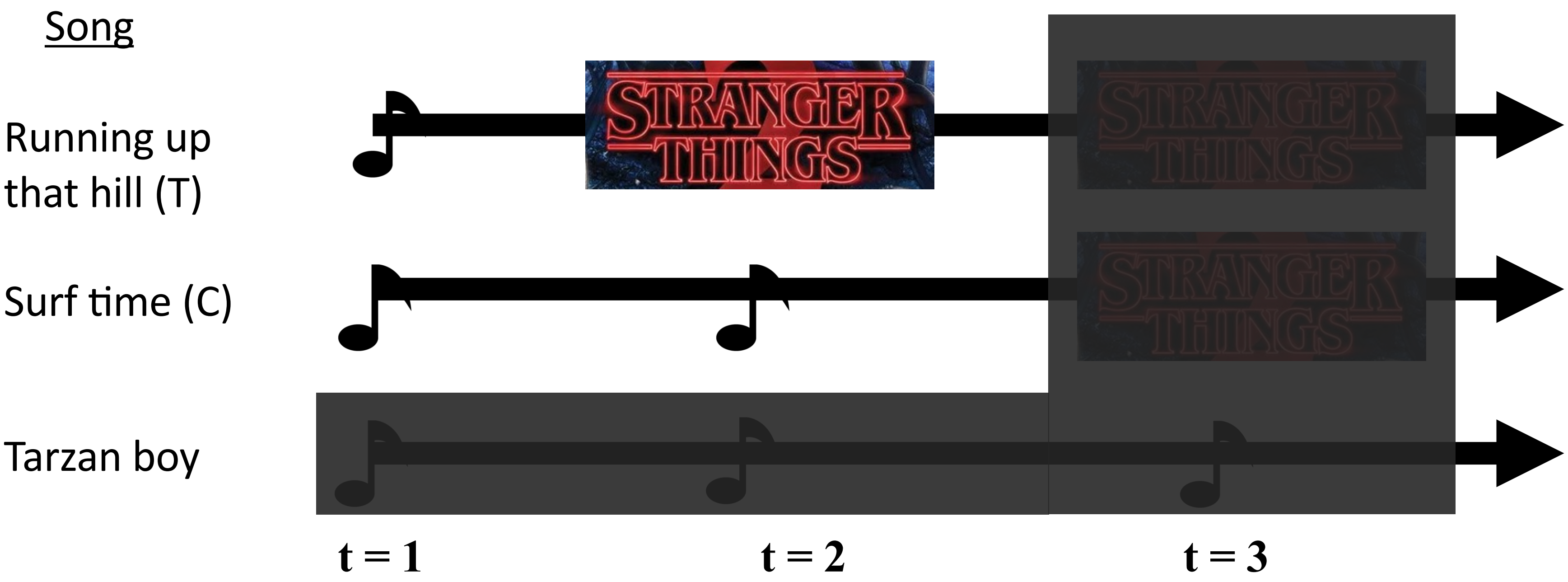
***The difference in the changes = change in the treated – change in the control***
***= 3-1***
**= 2**

# DiD with Variation in Treatment Timing

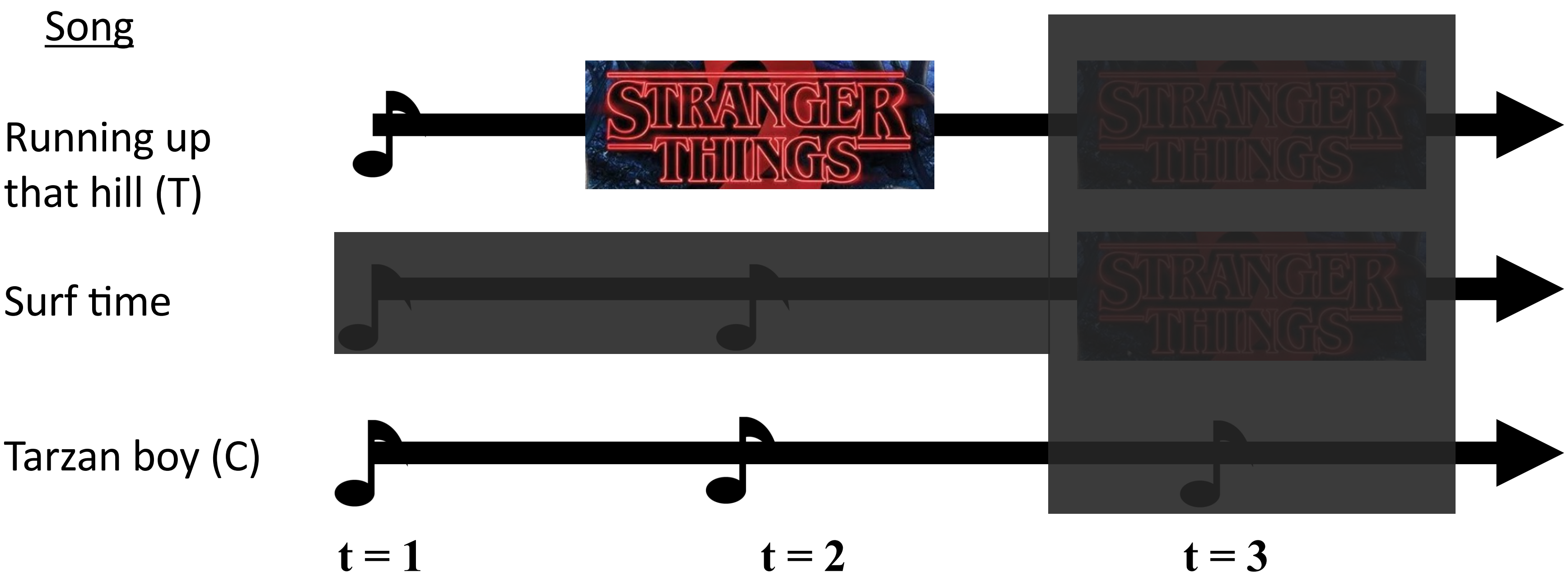



*There are different treatment cohorts: **early-treated, late-treated, and never-treated**.*
*Traditional DiD estimates a treatment effect as a **weighted average of two-group and two-period combination estimates**.*

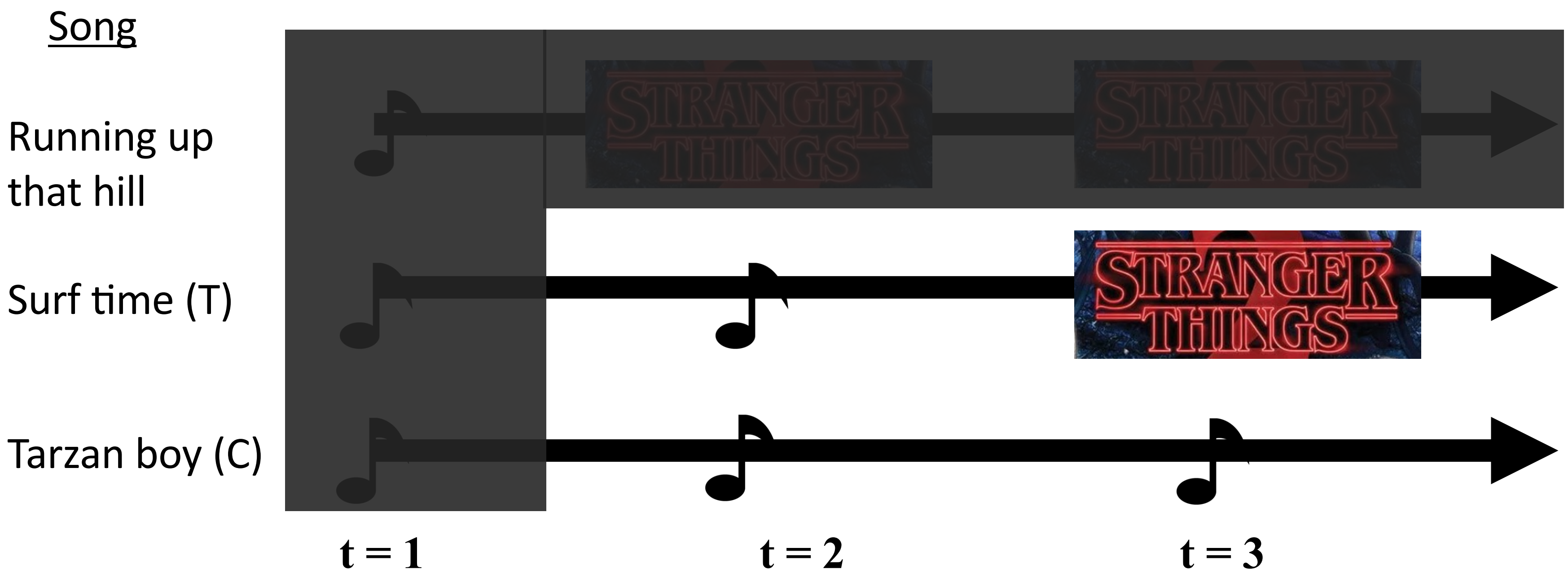
# DiD with Variation in Treatment Timing



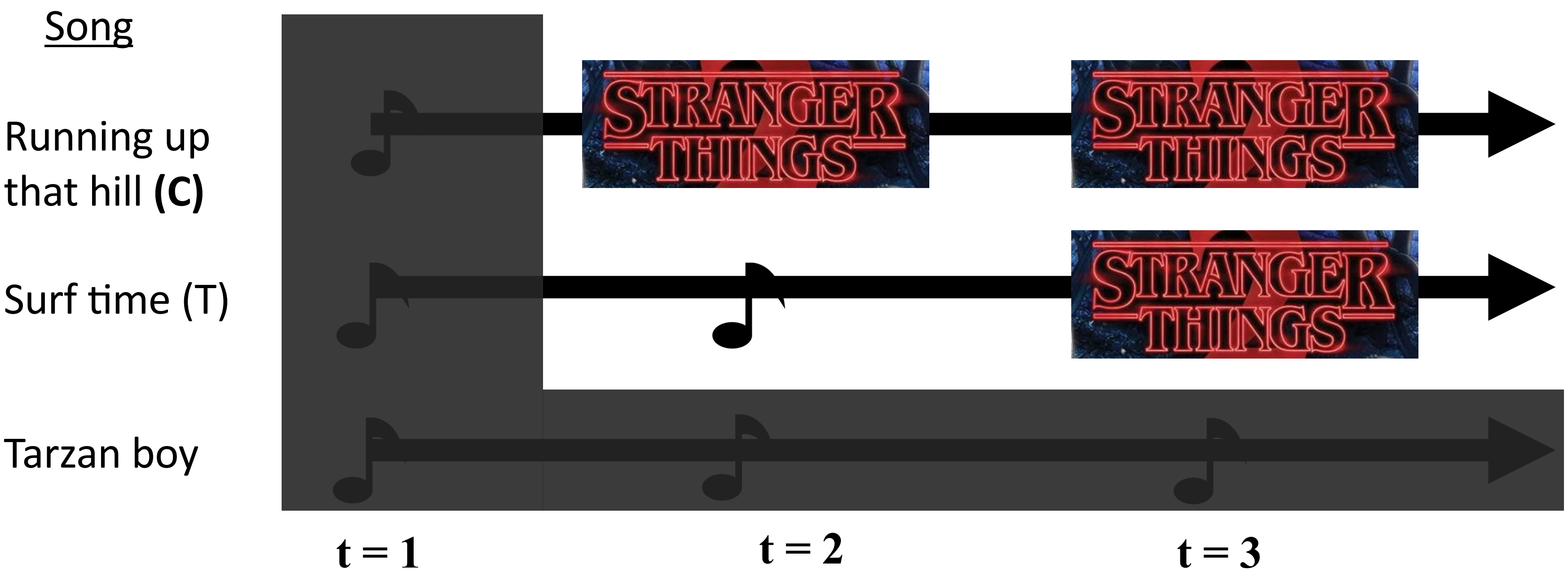
# DiD with Variation in Treatment Timing



# DiD with Variation in Treatment Timing

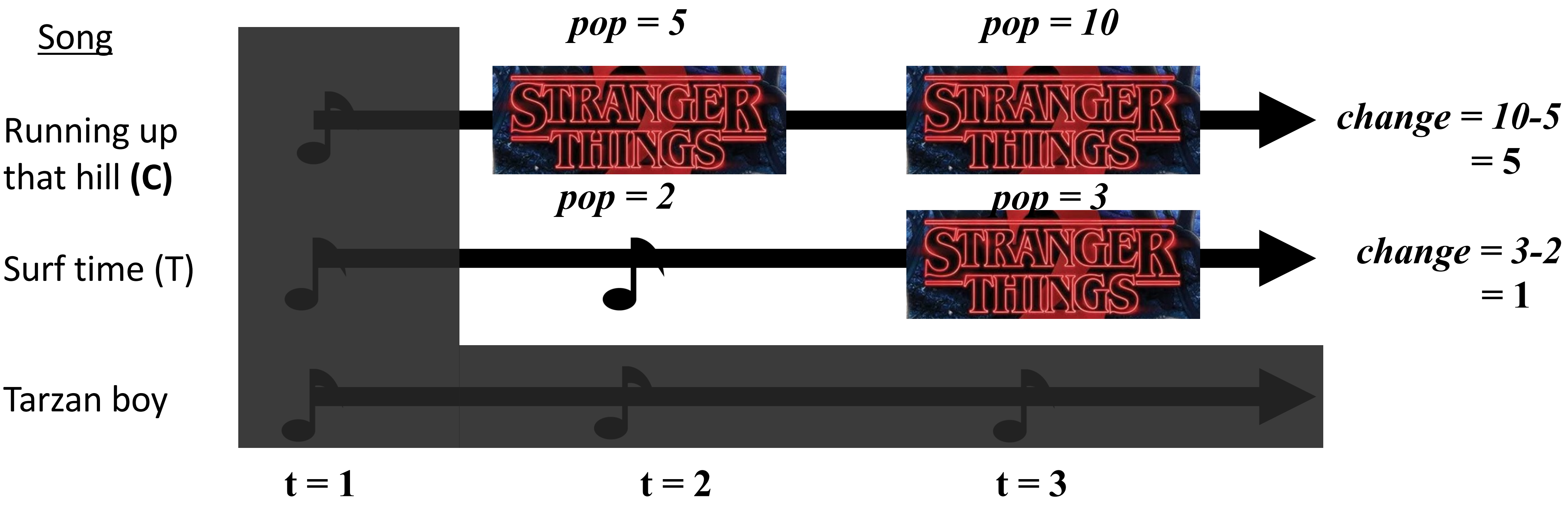


# DiD with Variation in Treatment Timing




***"Forbidden comparisons"***

# DiD with Variation in Treatment Timing




*The difference in the changes = change in the treated – change in the control*
*= 1-5*
**= -4**

# How to Address This Issue

- Traditional DiD estimator can provide misleading estimates when there is variation in treatment timing
- We can use DiD making sure to avoid "forbidden comparisons”
- Use the clean controls, i.e., late- or never-treated by using staggered DiD methods
- Most used staggered DiD methods:
  - Staggered DiD by Callaway & Sant’Anna 2021
  - Stacked DiD by Cengiz et al. 2019

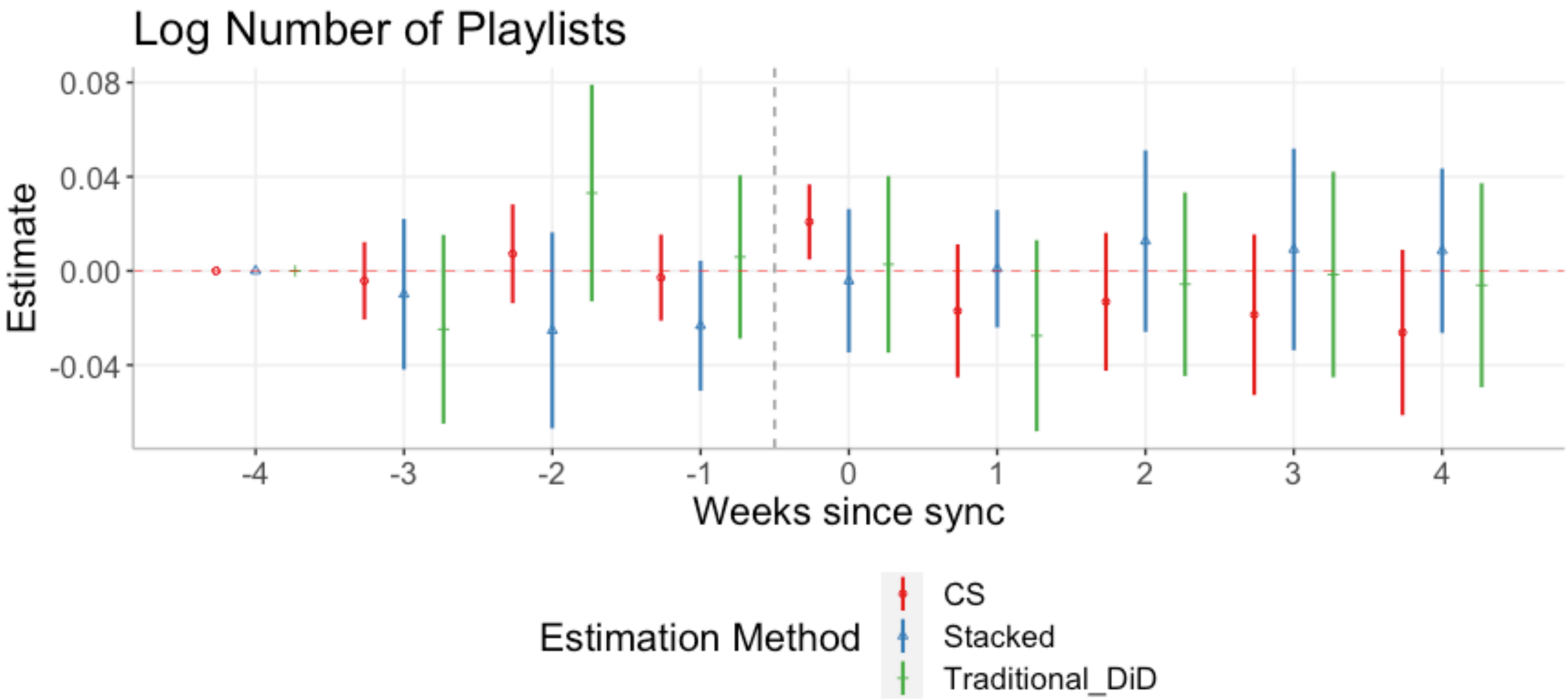
# Staggered DiD by Callaway & Sant'Anna 2021

- Practical Implementation
  - Estimates a group-time treatment effect
  - Aggregates estimated effects by group, exposure, or at an overall level
- Advantages
  - Available as an R package called "*did*"
  - Transparent selection of control group
  - Built-in options for customization (e.g., choosing the control group)
- Disadvantages
  - No covariate interactions to gauge heterogeneity
  - Can be computationally intensive

# Stacked DiD by Cengiz et al. 2019

- Practical Implementation
  - Create a ‘stacked dataset’ composed of group-specific
  - Run a modified DiD regression on the stacked data
- Advantages
  - Similar to traditional DiD in specification
  - Computationally inexpensive
  - Can include covariates to investigate heterogeneity
- Disadvantages
  - No dedicated package for implementation.

# Results


Log Number of Playlists
0.08
0.04
0.00
-0.04
Estimate
-4
-3
-2
-1
0
1
2
3
4
Weeks since sync
Estimation Method
CS
Stacked
Traditional_DiD

# Takeaways

- DiD is a powerful tool in situations where A/B tests are impractical
- However, if there is variation in treatment timing, the traditional DiD estimator might lead to misleading treatment effects
- We can avoid this by making sure to use the right controls with staggered DiD approaches

# Thank you for listening!

- 📧 Email:
  - nalagoz13@gmail.com
- 🔗 LinkedIn:
  - linkedin.com/in/nazli-m-alagoz/
- 📝 Medium:
  - medium.com/@nalagoz13

# Appendix

# Difference-in-Differences (DiD)

- *The notation*

$$Y_{it} = \alpha_i + \gamma_t + \beta^{dd} D_{it} + \epsilon_{it}$$

# Staggered DiD by Callaway & Sant'Anna 2021

- General
  - **Control Group**: Utilizes units treated late or never treated as control.
  - **Group-Time Effect**: Introduces the notion of average treatment effect over time for a specific cohort $ATT(g,t)$.
  - **Flexible**: Computes weighted averages of multiple treatment effects.
- Practical Implementation
  - Available as an R package called did.
- Advantages
  - 📦 Comprehensive R package.
  - 🔄 Allows for conditional Parallel Trends Assumption.
  - 🎯 Transparent selection of control group.
  - ✅ Implements multiple testing correction.
- Disadvantages
  - 🛠️ No covariate interactions to gauge heterogeneity.
  - 🕒 Computationally intensive.
  - 📊 Lower-level estimations can be imprecise.

# Stacked DiD by Cengiz et al. 2019

- General
  - **Cohort-Specific Datasets**: Units treated in the same period and their respective controls.
  - **Clean Controls**: Matches each treated unit with unaffected controls.
  - **Stacking**: Stacks the cohort-specific datasets.
  - **Fixed Effects**: Incorporates both unit-cohort and time-cohort categories.
- Practical Implementation
  - Stacked regression model for nuanced treatment effects.
  - Weighted averages for group-time average treatment effects on the treated.
- Advantages
  - 👍 Similar to traditional TWFE in specification.
  - ⚡ Computationally inexpensive.
  - 📊 Can include covariates to investigate heterogeneity.
- Disadvantages
  - 🛠️ No dedicated package for implementation.