TIME SERIES ANALYSIS

Quantitative Analysis

Week 12

REVIEW

- In the last class, we introduced **differences-in-differences** models, which are an approach to research design that is often used when you have **panel data** – i.e., observations of the same set of cases **before** and **after** the treatment whose effects you are interested in.
- This design is suitable when you have a **control group** – i.e. some set of observations of cases where the treatment / event did not apply.
- It is based on the **parallel trends assumption**, which is to say that we assume both the control and treated cases would have had similar trends if the treatment event had never happened.

REVIEW: PARALLEL TRENDS

- The historical example we looked at was John Snow's analysis to prove that sourcing drinking water from the heavily polluted River Thames was the cause of a fatal cholera outbreak in London in 1854.
- One borough – Lambeth – had changed its water source in 1852. Snow theorised that without this change, Lambeth's cholera cases would have followed **parallel trends** to the other boroughs of London.
- Using a differences in differences model, he showed that Lambeth's policy change had reduced cholera cases from the **predicted** number of cases that would have occurred with parallel trends.

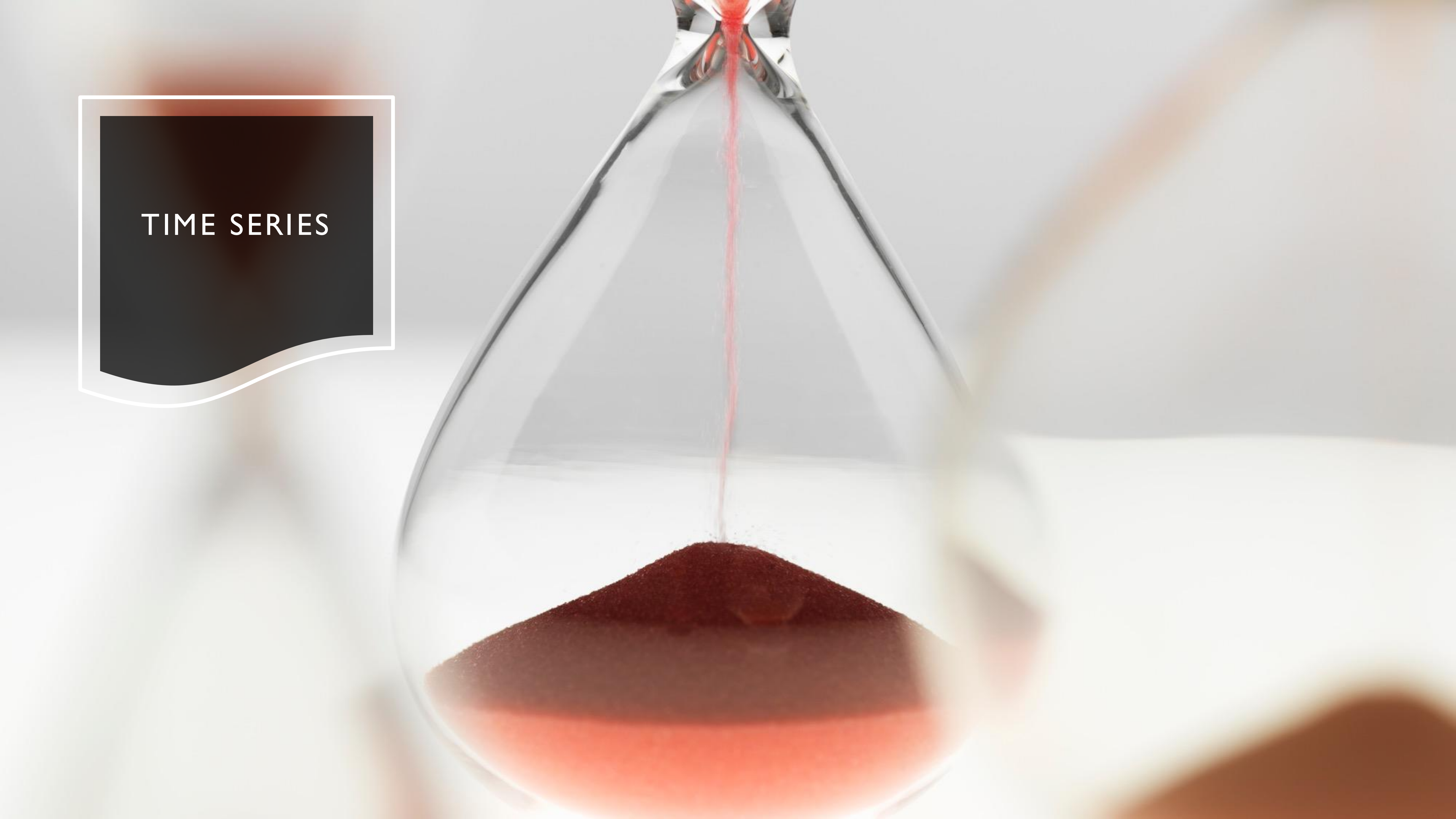
REVIEW: FIXED EFFECTS

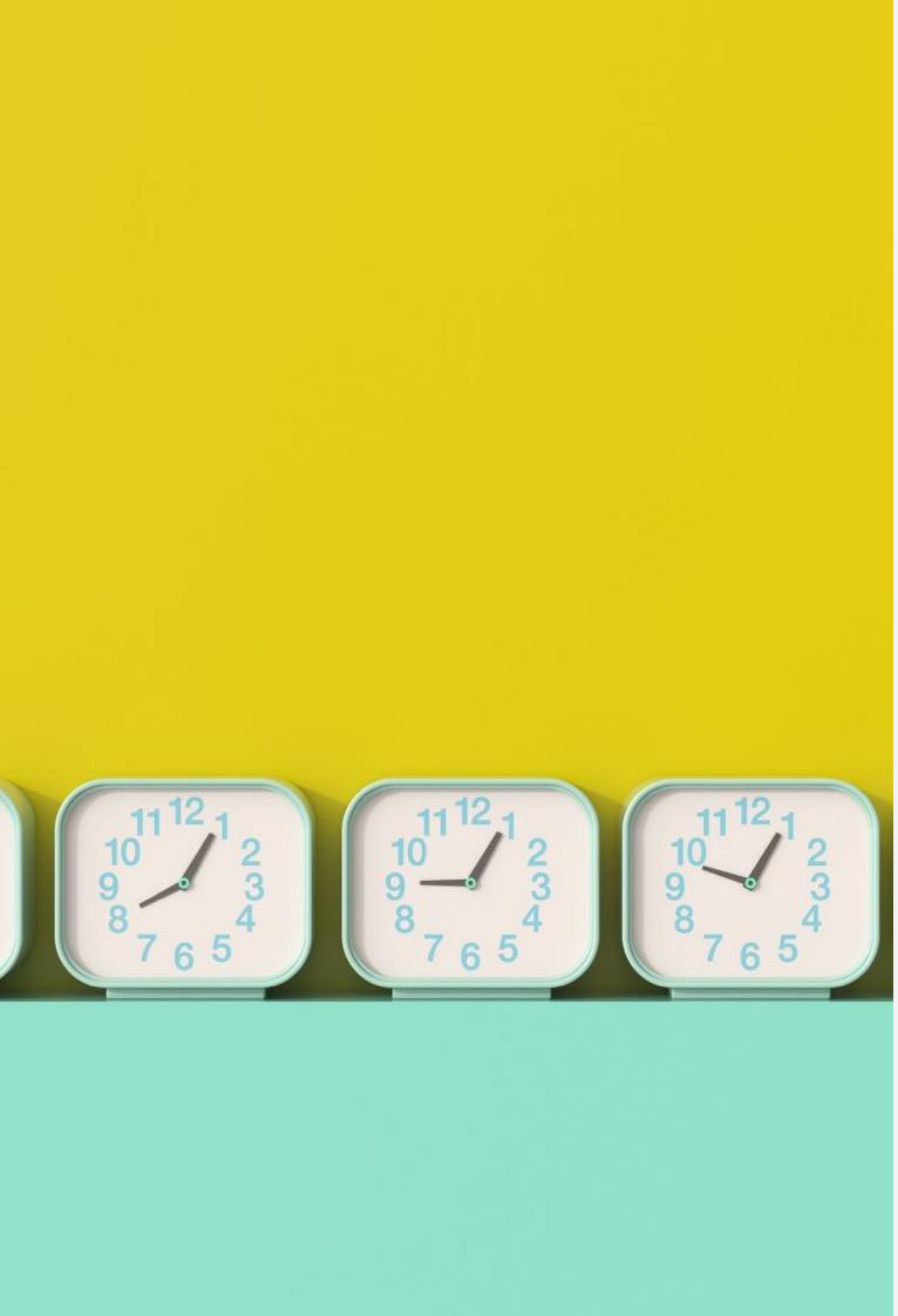
- We next moved on to looking at something called **fixed effects models.**
- Differences-in-Differences is actually a specific case of fixed effects. These models introduce dummy variables which allow for the fact that some of your cases will have certain parameters that are unobserved (which would usually cause **omitted variable bias**) but which remain **fixed** between observations.
- For example, if you are studying a group of countries over time, you know that there will be a lot of differences between Germany, France, and Italy – but you can argue that those **country-level effects** will remain fixed from year to year (there's no major shift in the unobserved characteristics of those countries over the few years you're observing).

DID VS. FIXED EFFECTS

- A very important distinction between DiD and Fixed Effects models is that for DiD, you must have a control group**.** You cannot use the parallel trends assumption unless you have cases where the treatment was not applied that you can use to calculate those trends!
- Standard Fixed Effects models do not require a control group. For example, you could apply a country-level fixed effect to a set of cases where every country had the treatment (e.g., they all experienced the same major event or signed the same treaty etc.).
- DiD is an attempt to replicate an experiment (with control and treatment groups) from observational data; Fixed Effects models are simply ways to improve estimations of complex cases.

TIME SERIES





REVIEW: TIME SERIES NOTATION

When we talk about time series data, we usually use the variable t to denote the time period we're looking at.

t or t_0 is the present, or the time period of interest (for example, the point where the event we're interested in studying happened).

We then denote earlier time periods by counting backwards in negative numbers: $t_{-1}, t_{-2}, t_{-3} \cdots t_{-n}$

Later time periods: $t_{+1}, t_{+2}, t_{+3} \cdots t_{+n}$

Variable X at time period t_{-3} would be written as X_{t-3}

REVIEW: PANEL DATA VS. TRUE TIME SERIES DATA

We used Differences-in-Differences (and Fixed Effects) models to study **panel data** – which is data from different time periods. How is this different from "real" time series data?

True **time series data**, is data that has a lot of observations at regular intervals.

- Daily stock closing prices;
- Monthly opinion polls;
- Annual demographic figures; etc.

Time series analysis provides us with methods for studying this kind of data by controlling for things like **seasonal cycles** and **long-term trends**, which are problems we don't encounter with cross-sectional / panel data.

SERIES COMPONENTS

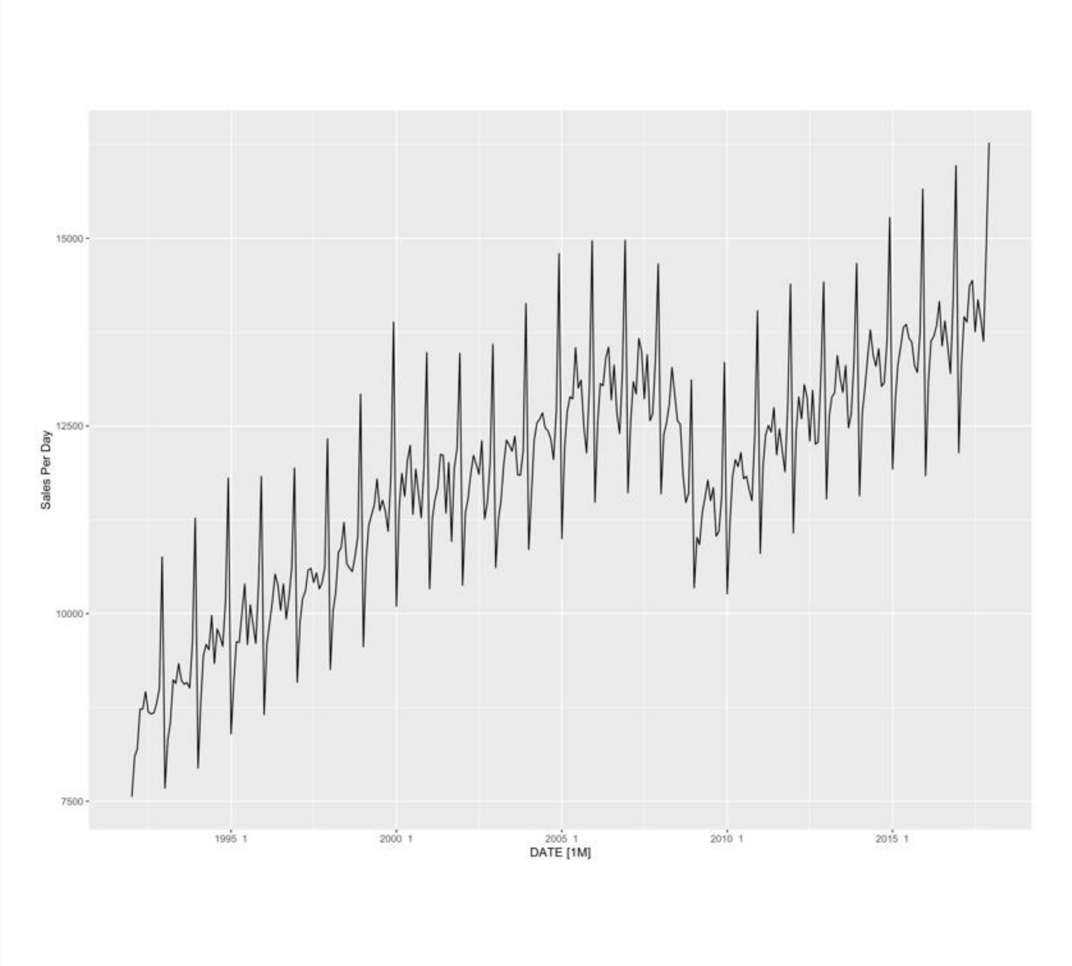
- A true time series – a set of observations over a long period of time – will be made up of three separate **components**. Each of these components contributes to the actual numbers we see in the series.
- The **trend** is the long-term movement of the variable.
  - For example, GDP has trended upwards in almost all countries over the past century.
- The **seasonality** is the regular, cyclical pattern of changes in the variable.
  - For example, temperatures and rainfall change on an annual pattern. Retail sales have both annual and weekly cycles. Online activity has a 24-hour cycle.
- Finally, the **remainder** is the part of the time series that is not explained by either the trend or the seasonality.
- <u>This is usually what we're interested in</u>, since it's the part of the data that would show responses to unusual events or outside influences.

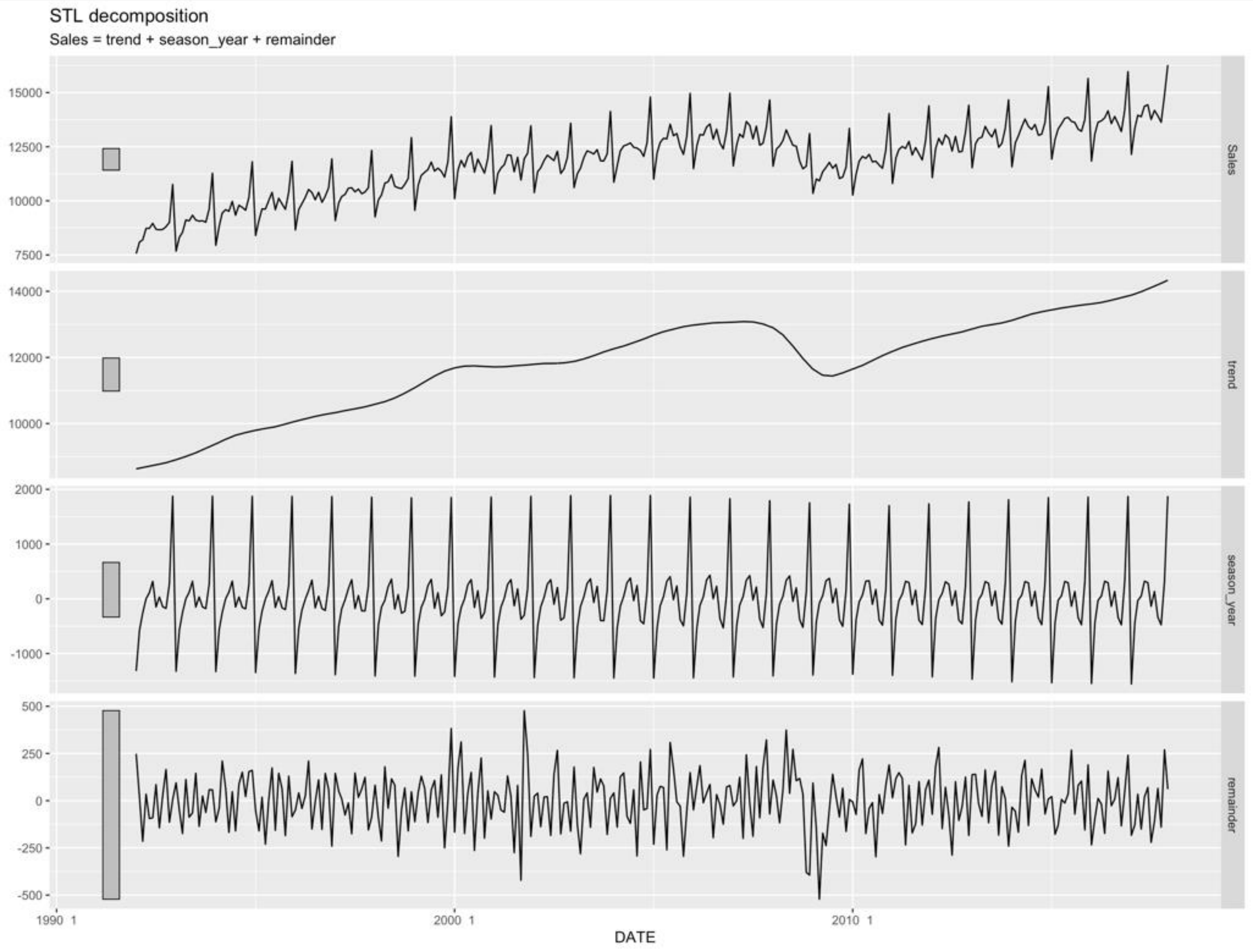
DECOMPOSITION

- Working with time series data generally involves **decomposing** the variable to discover and remove the components we don't want.
- The most simple way to remove the **trend** is to take the **first difference** of the variable: instead of the absolute value, we represent each period according to its change since the last period. This removes the long-term trend while preserving all of the detail in the data.
- This is an important step in many cases, because a lot of different measurements have risen or fallen over time, but many of them aren't related to each other at all. Leaving in the trend component would create **spurious correlations**.

SEASONALITY

- The seasonal cycle of many types of data is quite obvious and easy to see.
- This graph shows US retail sales on a month by month basis for thirty years. You can see the trend, of course, but the regular spikes on the graph also indicate an annual seasonality.
- The simplest way to remove seasonality is to figure out the **cycle length** (in this case, 12 months), then calculate the mean value for each point in the cycle (e.g., the mean for every January, every February, etc.), and subtract it from the observed values.
- In this way the seasonality is controlled and disappears from the graph.





SEASONAL AND TREND DECOMPOSITION WITH LOESS

- LOESS is a method for smoothing data and reducing noise, which is often used to plot smooth lines on graphs to make them easier to interpret.
- It's also very useful for time series analysis, because it allows us to quickly **decompose** all the elements of the time series, a process called STS.
- After decomposition, we're left with a **remainder** portion of the data. This is the variable we would most likely want to use in our regression analysis, now that the time series has been decomposed.


OVER TO R

FINAL PRESENTATIONS

- The objective of your group project is to produce a research presentation which you will present in class on January 22.
- This presentation should be **<u>five minutes long</u>**. This is pretty short!
  - If you have done additional analysis work that you want me to see, you can include extra slides at the end of your slide deck which you don't use in the presentation, but which I'll evaluate when grading.
- Since the time is short, you may nominate one or two group members to do the presentation.
- All slides MUST be submitted as PDF or PowerPoint files by midnight on **<u>January 21</u>**.

NEXT WEEK

- Next week's class will be run as a **workshop session**.
- The majority of the class will be an opportunity for groups to work together on their presentations for the following week (January 24).
- At the start of the class, I'll do a short example research presentation – showing you basically what your group should be aiming for with their presentation.
- I'll then give you a short while to confer together, after which every group will nominate one member to give a **brief** (~30 seconds) outline of their progress on the research project and any problems they are still dealing with.
- For the rest of the class you can work on your projects together and I'll consult with each group in turn. **Please make sure you sit with your group members at the start of the class!**