

# GEOG0149 Worksheet: Practical 5

---

## TODAY'S RESEARCH QUESTION

*Do high levels of homeownership increase the unemployment rate?*

## BACKGROUND

In recent decades, governments across the Global North have used a range of policy levers to try and increase homeownership rates. This thrust has been based on a belief that homeownership improves prosperity and wealth accumulation, fosters [good citizenship](#), helps build more stable communities, and is simply the [tenure of preference](#) for most people.

While there is much debate about the extent to which homeownership actually provides these personal and collective benefits, there has been surprisingly little consideration of the adverse impact that promoting owner-occupation could have on the labour market. This is in part because research on the economics of tenure dynamics has concentrated on [housing wealth and its role in welfare provision](#).

There is good reason to think that high rates of homeownership could adversely affect labour market performance. In a series of influential papers, Andrew Oswald showed that higher homeownership in an area or country is associated with a higher unemployment rate. He posited several mechanisms that could explain this pattern (we covered these in lecture 9 – they include the lower residential mobility of homeowners reducing job search and job matching, longer commutes by homeowners depressing productivity, reduced business formation in homeownership dominated areas etc) and provided correlational evidence for a positive association between homeownership and unemployment at the [level of American states and OECD countries](#). Other studies have corroborated these findings, with [Borg and Branden \(2018\)](#) arguing that the smaller size of labour markets with high homeownership rates reduces the efficiency of job matching, raising the probability of unemployment for residents regardless of their tenure (an externality effect). Interestingly though, at the micro-level research consistently finds that homeowners are less likely to be unemployed than tenants. So what's going on?

Today you are going to take a fresh look at this puzzle using recent data on European countries. We will begin by examining updating Oswald's work to see if there is still an association between higher homeownership and higher unemployment rates in the aftermath of the Global Economic Crisis. We will then examine change over time in both variables to try and get a better understanding of how and why homeownership and unemployment might be interlinked. Our overarching aim is to try and replicate [Oswald and Blanchflower's approach](#) using fresh data.

## TOPICS COVERED

1. Replicating Oswald's correlation analysis
2. Replicating Oswald's regressions

## PART 1: POINT IN TIME CORRELATIONS

- Load today's packages from library: dplyr, rgeos, rgdal, tmap and reshape2. You then need to import two spreadsheets: Unemp\_Europe.csv and Ownership\_Europe.csv. Call the new data frames 'Unemployment' and 'Ownership' respectively. They contain the following short time series of data (2010-2017) about 29 European countries obtained from [Eurostat](#):

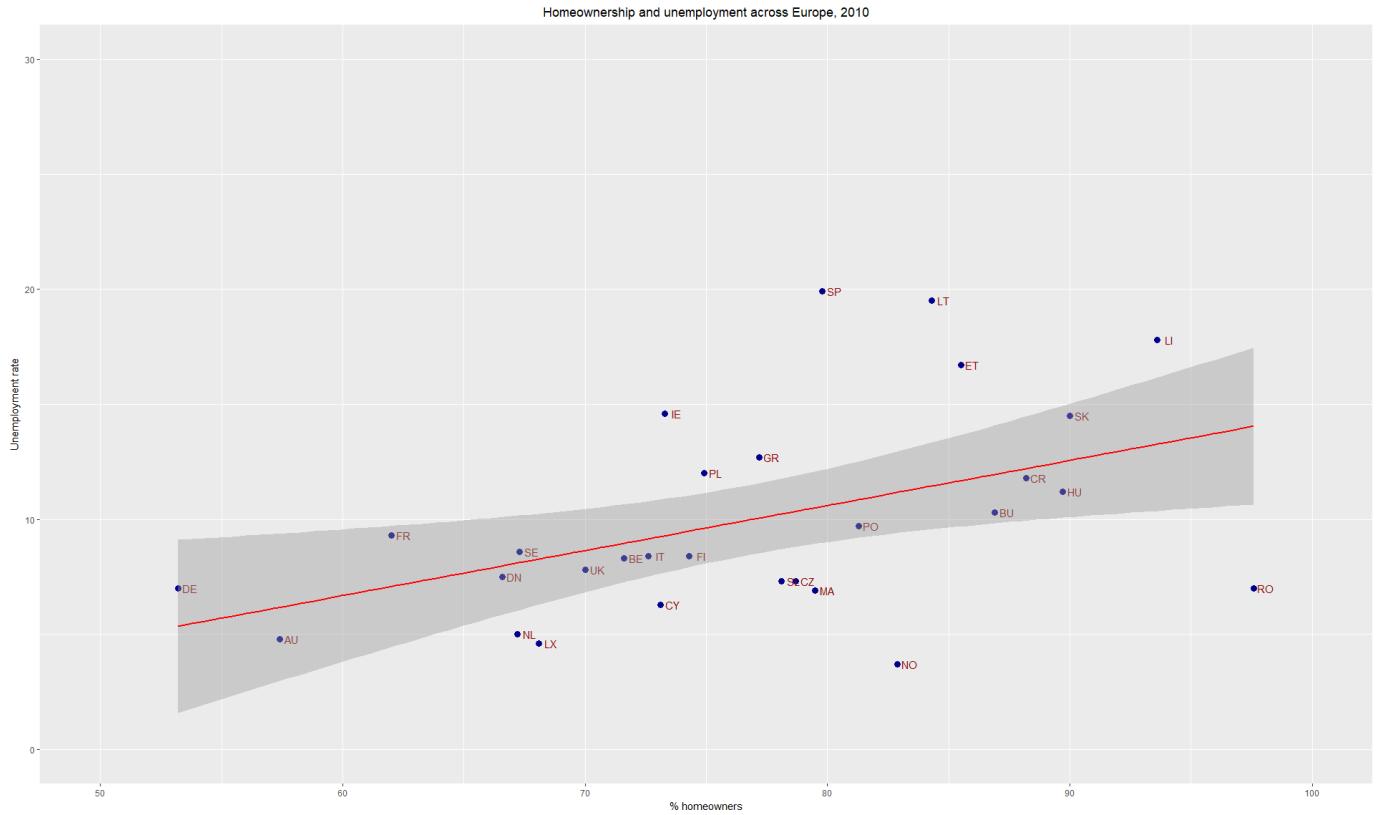
### 1. *Unemp\_Europe.csv*

Variable	Contents
Year	Year (2010 to 2017)
Country	Country name
Country_short	Abbreviated country name
Unemp_rate_total	Unemployment rate
Unemp_rate_u25	Unemployment rate among under 25s
Unemp_rate_2574	Unemployment rate among people aged 25-74

### 2. *Ownership\_Europe.csv*

Variable	Contents
Year	Year (2010 to 2017)
Country	Country name
perc_own	% of households who are owner-occupiers
perc_rent	% of households who rent

- Next merge the two data frames together into a new Europe\_rates dataset. To do this you'll need to key the merge on two identifier variables.
- Let's begin by testing Oswald's main argument by examining if higher rates of homeownership are associated with higher levels of overall unemployment across Europe (the Unemp\_rate\_total variable). To start off try extracting *only* data from 2010 and then (i) running an appropriate correlation test and (ii) producing a gg scatter plot with a regression line to illustrate the relationship between the variables of interest. This analysis mimics Figure 1 in [Oswald and Blanchflower's](#) working paper. Try to do this on your own and only use my code if you get stuck.
- You should end up with something like this, which you can see closely resembles their output. Clearly there is a weak but positive correlation between homeownership and overall unemployment in 2010.



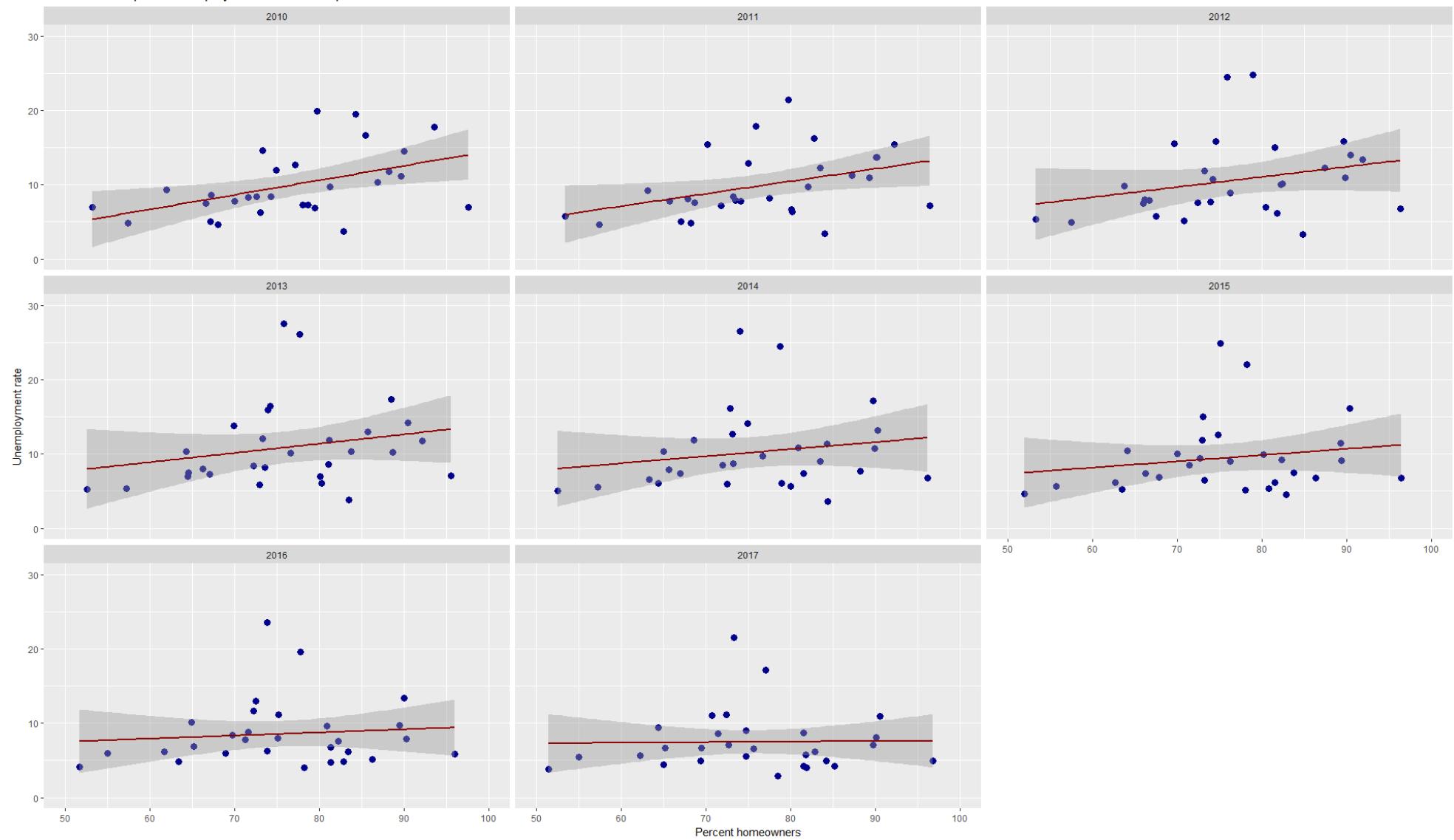
- Let's check whether this holds across the rest of the years. To do this first group the `Europe_rates` data frame and then use `summarise()` with `cor()` inside the brackets:

```
tests <- Europe_rates %>%
  group_by(Year) %>%
  summarise(Cor=cor(Unemp_rate_total,perc_own), N=n())
```

Year	Cor	N
1 2010	0.468	29
2 2011	0.399	29
3 2012	0.273	29
4 2013	0.231	29
5 2014	0.185	29
6 2015	0.179	29
7 2016	0.0990	29
8 2017	0.0199	29

- These Pearson tests are interesting as they suggest a weakening relationship over time between homeownership and aggregate unemployment rates across Europe. Not what Oswald has found! Let's look at this graphically by running separate gg scatter plots by year. Try this on your own. Hint: use `ggplot` to produce a plot object containing what you want to show then use `facet_wrap(~Year)` to produce multiple panels subset by year. It makes sense to switch off country labels for this exercise.

Homeownership and unemployment across Europe, 2010-2017



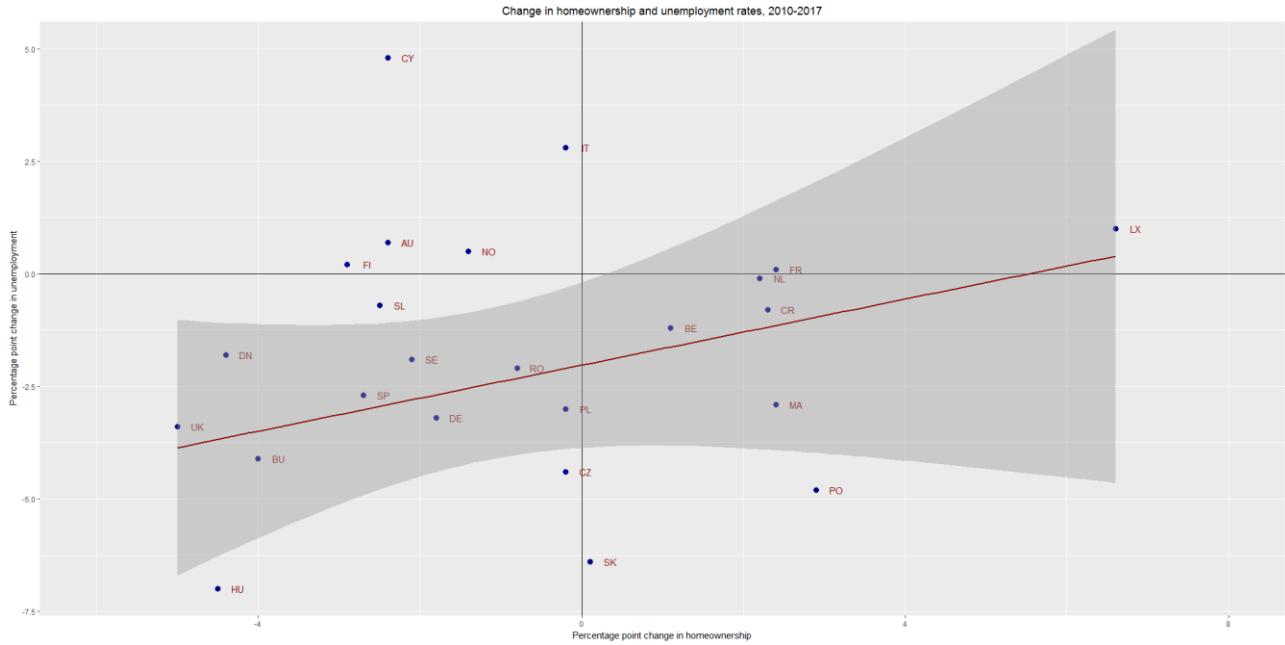
- Next try mapping the variables of interest using the Europe\_shapes file supplied on Moodle. To do this read in the shapefile to create a new object called Europe\_map. In order to do the merging we'll need to convert the rates dataset from long format (each row contains one country-year observation) to wide format (each row contains one country observation, with each column containing values on variables measured in particular years). This procedure is basically the reverse of what you did in Practical 2 when you created a person-year file — here we have a country-year file and are reshaping it back to ‘wide’ format.

```
# Reshaping code to convert from long (country-year) format to wide
Reshaped_rates <- reshape(Europe_rates,
                           direction = "wide", # The direction of the reshape
                           idvar = "Country", # The variable identifying each unit
                           timevar = "Year", # The panel (time) variable
                           v.names = c("Unemp_rate_total", # Time-varying variables
                                      "Unemp_rate_u25",
                                      "Unemp_rate_2574",
                                      "perc_own",
                                      "perc_rent"))
```

- Your new dataset should contain 29 rows (1 per country) and a lot of variables. These will hold the value of each variable in each of the years we are considering. So Unemp\_rate\_total.2010 contains the unemployment rate in 2010 while Unemp\_rate\_total.2017 holds the 2017 rate information and so on. As the data frame only contains one row per country it can now be merged onto the shapefile for mapping. Try mapping the 2017 unemployment and homeownership rates using tmap.

## PART 2: COMPARING CHANGE OVER TIME

- So far you've updated the first part of Oswald and Blanchflower's work by examining point-in-time correlations. Let's now move on to replicate their Figure 2 for European countries after 2010. This requires us to compare *change over time* in x (homeownership) to *change over time* in y (unemployment). If Oswald's hypothesis holds then the two should be positively associated. Compute the change variables now using the Reshaped\_rates wide format dataset by subtracting 2017 values from 2010 values on both our key variables.
- Now run an appropriate correlation test and produce a scatterplot of the change scores.



- You can see that during this period unemployment fell in most countries. Homeownership also fell in many, but by no means all, European countries. Overall there is a very weak relationship between the change rates. You can also try running linear regressions to see how these compare with the results of Oswald and Blanchflower's Figure 2.

### PART 3: MODELLING UNEMPLOYMENT

- Your final practical task is to have a go at creating a scaled down version of Oswald and Blanchflower's time series models predicting unemployment rates using (i) lagged homeownership rates (i.e. homeownership levels *in previous years*) and (ii) a suite of controls as independent variables. We are going to include measures of median age (called Median\_age) and % with tertiary education (called perc\_degree) as our controls as these are readily accessible from Eurostat. I've prepared files containing these variables for you: they are called Age\_Europe.csv and Education\_Europe.csv respectively.
- Read in the control variable spreadsheets and merge them onto the Europe\_Rates master dataset you already have in R. We then just need to generate a couple of new variables. First, let's follow O & B and set our dependent variable to be the *log of unemployment in year t*. This is to reduce the skewness you'll see if you create a histogram of it. Do this now and also use the Shapiro-Wilk's test (`shapiro.test()`) to test the null hypothesis that the unemployment variable's distribution does not differ from a normal distribution. After that you can compute a new variable called log\_unemp to hold the natural logarithm of unemployment (use the `log()` function to do this). Does this new variable seem more normally distributed?

- The last thing we must do is compute the *lagged homeownership and unemployment rates* for use in the model. These variables just measure (i) homeownership and (ii) unemployment in the country *in the previous year*. We can do this using the `slide()` function from the DataCombine package encountered in Practical 2:

```
Europe_rates <- Europe_rates %>% arrange(Country, Year) # For easier visualizing

Europe_rates <- slide(Europe_rates, Var ='perc_own', # Select variable to slide
                      TimeVar = 'Year', # Time variable
                      NewVar= 'perc_own_lag', # New variable's name
                      GroupVar = 'Country', # Group indicator
                      slideBy = -1)
```

- Now fit linear regressions predicting `log_unemp`. Build up your model gradually by including (i) only lagged homeownership; (ii) plus age and education; (iii) plus year as either a linear term or factor variable (you choose!) and finally (iv) plus country dummy variables to act as fixed effects picking up everything else that could differ across states. We won't include lagged unemployment in our model as O & B do because of the risk of temporally autocorrelated residuals.
- When you've finished export and interpret your models. What do they tell us about the Oswald hypothesis? You can see if the last model could be improved if you want – e.g. by centring predictors, testing interactions (this could be especially interesting!), examining diagnostics, producing predictions plots and/or by mapping residuals. These techniques have been described earlier in this course or in earlier modules so look back at your notes if you need to.

	base	sociodem	years	countries
(Intercept)	1.298*** (0.235)	0.913 (0.898)	126.257*** (34.264)	51.238 (29.098)
perc_own_lag	0.011*** (0.003)	0.012** (0.003)	0.014*** (0.003)	0.050*** (0.010)
Median_age		0.009 (0.017)	0.034 (0.018)	-0.118** (0.041)
perc_degree		0.000 (0.005)	0.008 (0.006)	0.017 (0.010)
Year			-0.063*** (0.017)	-0.024 (0.015)
R-squared	0.062	0.064	0.123	0.901
adj. R-squared	0.058	0.050	0.105	0.882
sigma	0.449	0.451	0.438	0.159
F	13.367	4.525	6.953	48.240
p	0.000	0.004	0.000	0.000
Log-likelihood	-124.661	-124.498	-117.856	103.329
Deviance	40.590	40.525	37.958	4.294
AIC	255.322	258.996	247.711	-138.659
BIC	265.262	275.562	267.590	-26.010
N	203	203	203	203

---

## EVIDENCE BRIEF QUESTION

1. To what extent is a higher homeownership rate associated with increased youth unemployment across Europe?

*Hint: you already have the youth unemployment rate (youth here is defined as aged 16-24) and all the other variables you need to answer this question. When producing your response, think about why we might expect youth unemployment to be particularly strongly or weakly affected by the homeownership rate. To do this you'll need to consider the mechanisms Oswald argues could cause homeownership to raise unemployment – are these likely to be particularly relevant for younger adults? Why or why not?*