

Course > Section... > 4.2 Usi... > Ecologi...

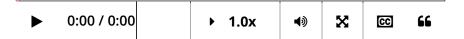
#### **Audit Access Expires Mar 24, 2020**

You lose all access to this course, including your progress, on Mar 24, 2020. Upgrade by Feb 18, 2020 to get unlimited access to the course as long as it exists on the site. **Upgrade now** 

# Ecological Fallacy Ecological Fallacy

Start of transcript. Skip to the end.





#### RAFAEL IRIZARRY: Throughout this section,

we have been comparing regions of the world.

We have seen that on average some regions do better

than others in health outcomes and economic outcomes.

Video



#### Download video file

## **Transcripts**

<u>Download SubRip (.srt) file</u> <u>Download Text (.txt) file</u>

## **Textbook link**

This video corresponds to the <u>textbook section on the ecological fallacy</u>.

## **Key points**

- The breaks argument allows us to set the location of the axis labels and tick marks.
- The *logistic* or *logit transformation* is defined as  $f(p) = \log \frac{p}{1-p}$ , or the log of odds. This scale is useful for highlighting differences near 0 or near 1 and converts fold changes into constant increases.
- The *ecological fallacy* is assuming that conclusions made from the average of a group apply to all members of that group.

## Code

```
# define gapminder
library(tidyverse)
library(dslabs)
data(gapminder)
# add additional cases
gapminder <- gapminder %>%
    mutate(group = case when(
        .$region %in% west ~ "The West",
        .$region %in% "Northern Africa" ~ "Northern Africa",
        .$region %in% c("Eastern Asia", "South-Eastern Asia") ~ "East A
        .$region == "Southern Asia" ~ "Southern Asia",
        .$region %in% c("Central America", "South America", "Caribbean"
        .$continent == "Africa" & .$region != "Northern Africa" ~ "Sub-
        .$region %in% c("Melanesia", "Micronesia", "Polynesia") ~ "Paci
# define a data frame with group average income and average infant surv
surv_income <- gapminder %>%
    filter(year %in% present year & !is.na(gdp) & !is.na(infant mortali
    group by(group) %>%
    summarize(income = sum(gdp)/sum(population)/365,
                        infant survival rate = 1 - sum(infant mortality
surv income %>% arrange(income)
# plot infant survival versus income, with transformed axes
surv income %>% ggplot(aes(income, infant survival rate, label = group,
    scale_x_continuous(trans = "log2", limit = c(0.25, 150)) +
    scale_y_continuous(trans = "logit", limit = c(0.875, .9981),
                                       breaks = c(.85, .90, .95, .99, .
    geom label(size = 3, show.legend = FALSE)
```

Learn About Verified Certificates

© All Rights Reserved