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**

Comparing Distributions

Comparing Distributions

Start of transcript. Skip to the end.

0:00 / 0:00 1.0x

The exploratory data analysis we have conducted

has revealed two characteristics about average income distributions in 1970.

Using a histogram, we found a bimodal distribution

with the most relating to poor and rich countries

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Textbook link

This video corresponds to the textbook section on 1970 versus 2010 income distributions. Note that the boxplots are slightly different: the group variable in those plots was defined in section 10.7.1.

Key points

- Use `intersect()` to find the overlap between two vectors.
- To make boxplots where grouped variables are adjacaent, color the boxplot by a factor instead of faceting by that factor. This is a way to ease comparisons.
- The data suggest that the income gap between rich and poor countries has narrowed, not expanded.

Code: Histogram of income in West versus developing world, 1970 and 2010

```
# add dollars per day variable and define past year
gapminder <- gapminder %>%
    mutate(dollars_per_day = gdp/population/365)
past_year <- 1970

# define Western countries
west <- c("Western Europe", "Northern Europe", "Southern Europe", "Nort

# facet by West vs devloping
gapminder %>%
    filter(year == past_year & !is.na(gdp)) %>%
    mutate(group = ifelse(region %in% west, "West", "Developing")) %>%
    ggplot(aes(dollars_per_day)) +
    geom_histogram(binwidth = 1, color = "black") +
    scale_x_continuous(trans = "log2") +
    facet_grid(. ~ group)

# facet by West/developing and year
present_year <- 2010
gapminder %>%
    filter(year %in% c(past_year, present_year) & !is.na(gdp)) %>%
    mutate(group = ifelse(region %in% west, "West", "Developing")) %>%
    ggplot(aes(dollars_per_day)) +
    geom_histogram(binwidth = 1, color = "black") +
    scale_x_continuous(trans = "log2") +
    facet_grid(year ~ group)
```

Code: Income distribution of West versus developing world, only countries with data

```
# define countries that have data available in both years
country_list_1 <- gapminder %>%
    filter(year == past_year & !is.na(dollars_per_day)) %>% .$country
    country_list_2 <- gapminder %>%
    filter(year == present_year & !is.na(dollars_per_day)) %>% .$countr
    country_list <- intersect(country_list_1, country_list_2)

# make histogram including only countries with data available in both y
gapminder %>%
    filter(year %in% c(past_year, present_year) & country %in% country_
    mutate(group = ifelse(region %in% west, "West", "Developing")) %>%
    ggplot(aes(dollars_per_day)) +
    geom_histogram(binwidth = 1, color = "black") +
    scale_x_continuous(trans = "log2") +
    facet_grid(year ~ group)
```

Code: Boxplots of income in West versus developing world, 1970 and 2010

```
p <- gapminder %>%
    filter(year %in% c(past_year, present_year) & country %in% country_
    mutate(region = reorder(region, dollars_per_day, FUN = median)) %>%
    ggplot() +
    theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
    xlab("") + scale_y_continuous(trans = "log2")

 p + geom_boxplot(aes(region, dollars_per_day, fill = continent)) +
     facet_grid(year ~ .)

 # arrange matching boxplots next to each other, colored by year
 p + geom_boxplot(aes(region, dollars_per_day, fill = factor(year)))
```

Learn About Verified Certificates

© All Rights Reserved