

SPEC Lab R Resources: Data Visualization with `ggplot2` (part II): Group Work

Alix Ziff adapted from materials by Therese Anders

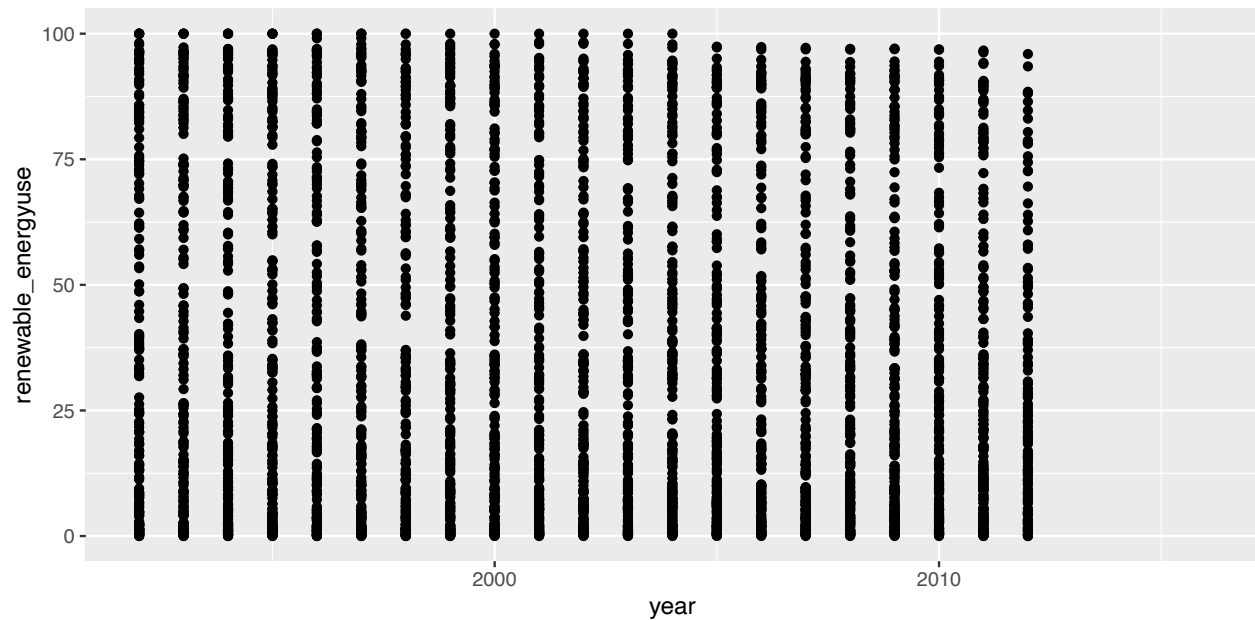
Summer 2020

`ggplot2` (continued)

#Summarizing patterns in scatterplots Let's use the same data from the Walk-Through-Work. Once you load your working directory and `wdi_cleaned_part2.csv` data, take a look.

```
library(ggplot2)
setwd("/Volumes/GoogleDrive/My Drive/Training Data Science/0. Training Data")
dat <- read.csv("wdi_cleaned_part2.csv")

ggplot(dat, aes(x = year, y = renewable_energysuse)) +
  geom_point()
```



```
table(dat$year)
```

```
##
## 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006
## 217 217 217 217 217 217 217 217 217 217 217 217 217 217 217
## 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
## 217 217 217 217 217 217 217 217 217 217
```

##Opacity **Exercise 1:** adjust the opacity of the plot points to help determine patterns in the data by

adjusting the values of alpha. What patterns can you deduce from the data?

##Jittering **Exercise 2:** Try to jitter the points. **A.** *Helpful Hint: set a seed to control the randomness.*

B. Change the default jitter value by adjusting the **width** and **height** parameters inside the `position_jitter ()` argument.

Trend Lines **Exercise 3:** Use the `stat_smooth ()` and `method = "lm"` arguments to overlay the scatterplot with a line of best fit of a linear model that regresses the proportion of renewable energy usage on the year.

Color Coding

Exercise 4: Create a binary variable that codes the wealth of countries to distinguish how the usage of renewable energies differs between richer and poorer countries.

A.: What generalizations can you make about renewable energy use 1. over time and 2. across wealth distributions.