





Explore data and R


Cleaning data


 **Video:** Cleaning up with the basics
7 min


 **Reading:** File-naming conventions
10 min


 **Reading:** More on R operators
10 min


 **Video:** Organize your data
6 min


 **Practice Quiz:** Hands-On Activity:
Cleaning data in R
1 question

 **Reading:** Optional: Manually create
a data frame
10 min

 **Video:** Transforming data
5 min

 **Reading:** Wide to long with tidyr
10 min

 **Ungraded Plugin:** Clean, organize,
and transform data with R
10 min

 **Practice Quiz:** Test your knowledge
on cleaning data
3 questions

Take a closer look at the data

Weekly challenge 3

Wide to long with tidyr

When organizing or tidying your data using R, you might need to convert wide data to long data or long to wide. Recall that this is what data looks like in a wide format spreadsheet:

	A	B	C	D	E	F	G	H	
1	Country Name	2010 [YR2010]	2011 [YR2011]	2012 [YR2012]	2013 [YR2013]	2014 [YR2014]	2015 [YR2015]	2016 [YR2016]	20
2	Antigua and Barb	88028	89253	90409	91516	92562	93566	94527	
3	Argentina	40788453	41261490	41733271	42202935	42669500	43131966	43590368	
4	Aruba	101669	102046	102560	103159	103774	104341	104872	
5	Bahamas, The	354942	359577	363584	367168	370633	374206	377931	
6	Barbados	282131	282987	283700	284296	284825	285324	285796	
7	Belize	322464	330237	338000	345715	353366	360933	368400	

Wide data has observations across several columns. Each column contains data from a different condition of the variable. In this example the columns are different years.

Now check out the same data in a long format:

	A	B	C
1	Country Name	Year	Population
2	Antigua and Barb	2010	88028
3	Antigua and Barb	2011	89253
4	Antigua and Barb	2012	90409
5	Antigua and Barb	2013	91516
6	Antigua and Barb	2014	92562
7	Antigua and Barb	2015	93566
8	Antigua and Barb	2016	94527
9	Antigua and Barb	2017	95426
10	Antigua and Barb	2018	96286
11	Antigua and Barb	2019	97118
12	Argentina	2010	40788453

To review what you already learned about the difference, **long data** has all the observations in a single column, and the variable conditions are placed into separate rows.

The pivot_longer and pivot_wider functions








There are compelling reasons to use both formats. But as an analyst, it is important to know how to tidy data when you need to. In R, you may have a data frame in a wide format that has several variables and conditions for each variable. It might feel a bit messy.

That’s where pivot_longer()comes in. As part of the tidyr package, you can use this R function to lengthen the data in a data frame by increasing the number of rows and decreasing the number of columns. Similarly, if you want to convert your data to have more columns and fewer rows, you would use the pivot_wider() function.

Additional resources

To learn more about these two functions and how to apply them in your R programming, check out these resources:

- **Pivoting** : Consider this a starting point for tidying data through wide and long conversions. This web page is taken directly from tidyr package information at **tidyverse.org** . It explores the components of the pivot_longer and pivot_wider functions using specific details, examples, and definitions.
- **CleanItUp 5: R-Ladies Sydney: Wide to Long to Wide to...PIVOT** : This resource gives you additional details about the pivot_longer and pivot_wider functions. The examples provided use interesting datasets to illustrate how to convert data from wide to long and back to wide.
- **Plotting multiple variables**   This resource explains how to visualize wide and long data, with ggplot2 to help tidy it. The focus is on using pivot_longer to restructure data and make similar plots of a number of variables at once. You can apply what you learn from the other resources here for a broader understanding of the pivot functions.

Mark as completed