

Note to grader:

Please answer each question in isolation. For example, if during survey 2, someone had a great answer that included lots of detail, but then gave a lackluster response to survey 3, that is ok. Some learners also seemed to respond to the wrong question. These have been moved and noted accordingly.

Final scores are calculated by dividing actual points by total possible points for each question.

Question 1: What changes, if any, would you make to this dataset in order to do analysis or make a visualization?

country	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910
Afghanistan	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2 Albania	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3 Algeria	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4 Andorra	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5 Angola	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
6 Antigua and Barbuda	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
7 Argentina	2070	2490	2820.0	2860	3.80e+03	3990	6.26e+03	6.26e+03	7.62e+03	5.94e+03	8.91e+03
8 Armenia	131	135	130.0	127	1.42e+02	126	1.44e+02	1.61e+02	1.62e+02	1.72e+02	1.68e+02
9 Australia	10200	11400	11400.0	11200	1.16e+04	12100	1.44e+04	1.55e+04	1.68e+04	1.46e+04	1.75e+04
10 Austria	27700	28400	25700.0	25600	2.69e+04	28100	3.36e+04	4.22e+04	5.90e+04	4.22e+04	5.76e+04
11 Azerbaijan	1270	1300	1260.0	1230	1.37e+03	1220	1.39e+03	1.56e+03	1.57e+03	1.66e+03	1.62e+03
12 Bahamas	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
13 Bahrain	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14 Bangladesh	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15 Barbados	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
16 Belarus	1970	2020	1950.0	1910	2.12e+03	1900	2.16e+03	2.42e+03	2.44e+03	2.57e+03	2.52e+03
17 Belgium	49700	46700	48900.0	52700	5.04e+04	50300	5.75e+04	5.84e+04	5.79e+04	5.78e+04	6.13e+04
18 Belize	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
19 Benin	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
20 Bhutan	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
21 Bolivia	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
22 Bosnia and Herzegovina	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
23 Botswana	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

	Possible points	Actual points
"Not sure" or equivalent? Zero here and across all cells except last	0	_____
Pivot data	5	_____
Specifically mention pivot_longer or an action that would be pivoting into a longer format	2	_____
Handle missing values	2	_____
Clean names and/or special characters in country names	1	_____
Consistent formatting of numeric data	1	_____
Standardize colnames so they aren't year numbers	1	_____
Transform data values	1	_____
Check for duplicated values	1	_____
Discretionary - anything else you liked?	1	_____
Total points	15	_____

Question 2: In your own words, describe practices you might use to make your work more transparent and reproducible in R.

	Possible points	Actual points
"Not sure" or equivalent? Zero here and across all cells except last	0	_____
Use scripts or files of some kind (not console)	5	_____
Version control	3	_____
Share code	3	_____
Use Rmd or equivalent	3	_____
Commented code and/or explained what you did	3	_____
Share data / data dictionary (where ethical)	2	_____
set.seed()	1	_____
Use relative paths	1	_____
Use functions for repeated code	1	_____
Unit tests or have someone else review the code	1	_____
Discretionary - anything else you liked?	1	_____
Total points	24	_____

Question 3: In your own words, describe a process you would typically take when investigating a new dataset.

	Possible points	Actual points
"Not sure" or equivalent? Zero here and across all cells except last	0	_____
Inspect the data (View, head, tail, glimpse, or described equivalent) - just dimensions can get (2.5)	5	_____
Check for missing data	3	_____
Check for data that needs recoding	3	_____
Sort data / look for max/min or outlying values / summary stats	2	_____
Check variable types/class	2	_____
Check column names	2	_____
Check for problems importing the data	1	_____
Check for duplicates or do some kind of grouping	1	_____
Do a quick visualization such as a histogram	1	_____
Discretionary - anything else you liked?	1	_____
Total points	21	_____