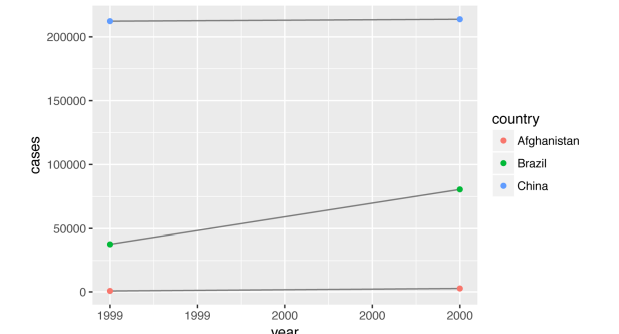
**RATHNAVEL SUBRAMANIAM COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS) SULUR, COIMBATORE.**

**ASSIGNMENT QUESTIONS**

**UNIT IV**

1. Re-create the plot showing change in cases over time using table2 instead of table1. What do you need to do first?



1. Why does this code fail?

table4a %>%

**pivot\_longer**(**c**(1999, 2000), names\_to = "year", values\_to = "cases")

*#> Error: Can't subset columns that don't exist.*

*#> ✖ Locations 1999 and 2000 don't exist.*

*#> ℹ There are only 3 columns.*

1. Compute the rate for table2, and table4a + table4b. You will need to perform four operations:

* Extract the number of TB cases per country per year.
* Extract the matching population per country per year.
* Divide cases by population, and multiply by 10000.
* Store back in the appropriate place.

Which representation is easiest to work with? Which is hardest? Why

gather() has a names\_ptype argument, e.g. names\_ptype = list(year = double()). What does it do

1. a. Add a surrogate key to flights

b. What does anti\_join(flights, airports, by = c("dest" = "faa")) tell you?

c. What does anti\_join(airports, flights, by = c("faa" = "dest")) tell you?

1. Perform the following
2. Compute the average delay by destination, then join on the air ports data frame so you can show the spatial distribution of delays.
3. Add the location of the origin and destination (i.e. the lat and lon) to flights.