* 2nd and 3rd parts aren’t rocket science
* Star schemas are really helpful for dashboard where accessing information needs to be super quick….queries that take 2-3 hours to generate are really only okay for reports
  + Star Schemas make it fast be precomputation…you don’t have to do GROUP BY’s when querying the fact table
* A summary table has all pre-joins with no dimension tables!
* For the third part, he wants us to make some graphs
  + Text

    Description automatically generated
  + For graphs, you can use plot (use line graph) but should use ggplot
* We will be graded on the fact that our fact tables works
* R has a dashboard engine called Shiny
* “Fact tables are a bear to build and flipping big” – Martin
* Should make two fact tables to look at articles by year/quarter and then by author
* Dimension tables are *very* useful because they allow you to populate drop-down menu items; according to Martin…that is the real reason why you have dimension tables
* Approach to building: Fact BY Dimension
* For dimensions where the selection is “ALL”, he is using PK 0
* An example author fact table
  + A picture containing text

    Description automatically generated
* To computer and populate a fact table, you use your program of choice and “do what you gotta do”
  + Text

    Description automatically generated
    - dbSend uses 0 memory but is slow
    - TODO: google dbBind()
* Can build your fact table in a dataframe and then load all at once
  + dbWriteTable is fast but uses lots of memory
* This will be “ugly, grindy, not exciting code” to do computations
* Martin recently worked on a project where he did dbSend followed by dbBind ….look at this
* Looks like we should use paste0
  + Graphical user interface, text, application

    Description automatically generated

Possible Data Mining from Lecture

* Is our editorial review process too cumbersome; do our articles take a while after submission to be published
  + e.g. daysToPublication by Year by Quarter
    - Text

      Description automatically generated