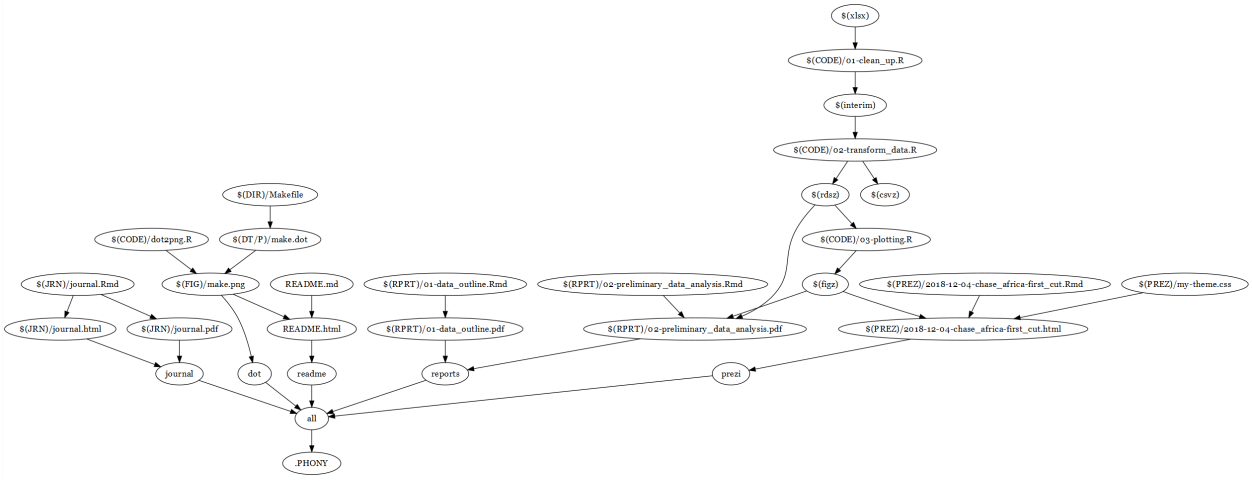
Journal–Chase Africa

Contents

1 Makefile

This is what my makefile looks like graphically at the minute:



2 Monday 29.10.2018

1. Initialise repository
2. Move data into `raw` data folder, make sure not in repository.
3. What are the data files? Outline sheets

3 Tuesday 30.10.2018

1. Finish the CHAT file outline
2. Go through all files and complete outline of data. write up in `01-data_outline.Rmd`
3. Update makefile

4 Thursday 22.11.2018

1. OK, now I'm meant to do some 'analysis'.
2. OK, so first thing is consolidating all the data into one file. Doing this manually will be a pain, but also really bad practice. Let's see if `readxl` might not be an option.
3. OK, Dandelion 2014 done.
4. Dandelion 2015 - mixed data in one of the date cells... A bitch to disentangle programmatically, ha.
5. OK, dandelion 2016. Had to remove a whole "funding period", not sure how this works!
6. Tried to dabble with moving averages, but got a bit too intense..
7. OK, Dandelion 2017

5 Friday 23.11.2018

1. Hmm, should I actually keep the odd ones, the weird funders? I mean in the end they will presumably be still in the analysis, so I might as well sort this out now...
2. OK, back to 2016 and add "amplify change" back in.
3. And back to 2017 with Amboseli added in.
4. 2018 now. Found error in pills CYP in the final sheet.
5. OK, all data together now. Start row bind with 2018 dataframe, just to keep the columns in that order, since there are most in that table.
6. Now sort by date.
7. Now update makefile. I really need to start doing this earlier, not later!
8. Update readme
9. TODO:

- prepare presentation folder/gh_pages
- group data monthly?

6 Monday 26.11.2018

1. Preapre presentation folder/gh_pages.. xaringan, right?
2. ALso update readme and add to makefile.
3. OK, now what to do with the data. Should I have a look at another dataset? No, keep it simple and do dandelion only.
4. Start preliminary analysis report.
5. Rename variables so they are fp_ family planning, ihs for integrated health services and fund for funding related stuff.
6. Derive totals for family planning and check with the ones in the excel spreadsheet - correct.
7. Derive CYP totals

7 Tuesday 27.11.2018

1. Fix NAs e.g. in 2015 pills.
2. Makefile: interim data and transform data script.
3. OK, derive CYP total, as well as sub totals. Compare with original version. 71 are the same and the other 81 are not.. But that's not my error, it's theirs since they had teh CYP conversion factors entered in wrong.
4. There are two more totals, IHS and IHS+FP together, do those and double check. In order to do those need to rename the `ihs_`variables. And careful to not count positive tests as people.. One check revealed swapped columns in 2018. other check revealed: in 2017 the total IHC formula included the hiv pozitives.
5. Now figure out smoothing lines - I guess in ggplot if I want to use gganimate..
6. But maybe first add summaries per funding round. OK
7. No, even before, add the list of variables to the appendix. OK
8. Now update makefile. OK
9. Now back to summaries per year and per round. How do you do this: different groups of variables need different funcitons: some sum, some mean, some max, some first(). OK, what di I realy want to do:

First by funding round:

- drop date and venue
- sum the next 28 variables.
- first the five funding variables - although one is the grouping one.
- sum all the remaining variables.

OK, so sum all non-funding variables. Do that using mutate, then summarise.

10. Now need cost per person and cost per Couple Year protection. OK, GBP per FP person checks out with the Excel data. And CYP as well - well, except for the CYPs that were wrong in the Excel files. And I assume the Ksh are correct as well, although only checked a few.
11. Now summarise by year and by funding round. Careful that the cost per person or per CYP is calculated correctly as well.
12. Only issue with summaries is that they include the non-standard funding rounds. Maybe I can have a look at what happens without them? Meh, actually, let's not complicate things. But I left the code in.

13. Now smoothing lines in ggplot2
14. check if kable can't have a bit more pzzaz? Done
15. OK< plots, set up makefile.
16. Start with gganimate, but it requries transformr apparently, and i'm on a train.
17. WHen I get to the plots, the zeros and missing are not great.. If it's an NA it shouldn't actuallly register as a zero. But that's what they have been entered in - albeit oddly, they look like dashes, but register as zeros?!

8 Wednesday 28.11.2018

1. Presentation stopped working there, but then started again... Phew, patience with gh_pages!
2. OK, somehow i don't know how to write a function for a ggplot chart - the variables are not passed ok!? But don't have time for that now.
3. Also with pdf figure size and dpi don't seem to be combinable as they are with html. So font size changed manually. Let's add a few more simple time series.
4. Also set fig.size globally
5. Also yesterday I used a trick to force floats to "H" bu Yihue, that said that the option to force "H" only works if you have a caption and have at least `out.extra` defined here.
6. OK, now aggregates, by round and by year.
7. Missing two round dates, well one really, made it up!
8. OK, barplot, now i need to gather the df, apparently that's the only way..
9. Then massive sidetracking trying to figure out how to have two discrete scales: one for the variable and the other to distinguish the fact that 2018 is incomplete. Which you can't do with alpha, since it only affects fill, not the border apparnelty..

9 Thursday 29.11.2018

1. Unrelated sidetrack: apparently gganimate and patchwork do not yet work together. There is an example on the wiki of using magic to combine gganimate gifs, so that's sth to look into
2. Write some notes under blog ideas about how gganimate works. OK, I think I have the gist of it..
3. See if I know how to animate two lines on one chart, one after another? Prob not.
4. Ugh, why is this so complicated. I now can't even do the thing where one line is drawn in the background and one added.. It worked fine in the prospective ageing talk? more repex time.. Christ, so it seems this works if you have separate dataframes!? FFS. And separate grouping variables. Got it. OK, that's half a blog post written.

10 Friday 30.11.2018

1. OK, let's get some gifs done then.
2. See how legend's might be added to the charts - and be inside the plot, not outside. OK, sorted, figured out how to do that for two separate layers as well. But doens't work with animations apparently!?

3. Repex of animation with legend. FFS, it's fine, just didn't show the extreme right in the rstudio viewer...
4. OK, so now first set of gifs seems fine.
5. Update makefile to make sure all is picked up smoothly.
6. OK, now can we transition from two lines to just one? I can use transition_layers maybe, but do two lines in the first layer, and only one in the second? Great, that works!
7. But can I get the legend to also disappear or rather be replaced with a new one? Prob not. Anyway, I can just overlay a new gif that starts with the black single line... shesh..
8. Can I transition between different `geom_smooth()`s? Probably not, unless I explicilty have it as a columns, that would probably be better?

11 Saturday 1.12.2018

1. not sure the makefile is working it seems to be going into repeating the plotting ones... let's try if this isn't because all the gifs are listed as targets, I'll try replacing that with just one.
2. OK, let's see if I can't do a bunch of loess regressions and cycle through them with state? But how will that work for SE? Hmm, is this necessary? But let's give it a try, what the heck. Maybe later actually... ANyway, i left my main work in the un-saved file on my other computer. That's a lesson learned right there!

12 Monday 3.12.2018

1. Loked into revealing the loess curves and there is a promising idea here but splits the line into segments and is kludgey... SO dropping that now..
2. I should probably add the sums as separate variables.. Or maybe no bother now. Not now.
3. Do double loess curve
4. Aggregate by funding round/period example.
5. Damn, really odd thing happening with a transition reveal, where the poitns are just jumping off and ahead of the lines!? i'll save as weird.gif, but have to move on now probably?
6. OK, al single and aggregate time series are done - oh, should have some bonus ones as well probably? I'll do those later.
7. Now need to do some rates stuff before the cost ones.
8. Change colours to chase colours.
9. Check spelling in document.
10. See if you can't get the CYP chart into an area one instead.
11. FUCK. why is there no 2018 data in the df_rounds table!? It's all good, the 2018 funding round was dated december 2017. So good to know: which date should we use?

12.1 Other ideas?

As has been noted, this preliminary analysis was only conducted on the Dandelion data. Using the data for all the sites will allow for comparison between them as well as mapping of the data if appropriate geographical data can be provided.

The CYPs seem most promising to base an analysis of impact on, but also have issues that need to be addressed e.g. the fact that all the CYPs are credited to the year the protection was delivered and not annualised over it's lifetime.

There weren't really enough data points here, but CYP cost could for example be modelled as a function of the structure of the contraceptive 'basket' e.g. ratio of long term to short term vs cost?