Mapping and analysing RV survey data

A personal analytical journey leading to the 2013 Atlas, and beyond

Daniel Ricard, DFO Gulf Region

December 17 2020

Background

Computing history

- 1980s: Basic on a Radio Shack computer (self-taught)
- 1990s: DOS, Scheme, C++, Linux, LaTeX (McGill, UBC, Simon Fraser)
- 2000s: SQL, R, postGIS, Perl (PRISMA, CSIRO, Dalhousie)
- 2010s: R package development, git, csasdown (Dalhousie, HBU, DFO)

Institutional context: as a student, you are on your own, you don't have a team to rely on, and you can't afford costly software

Ph.D. work at Dalhousie

- DFO colleagues at BIO and SABS (Branton, Smith, Clark, Shackell, ...)
- Participation in Georges Bank RV survey as a student
- ICES article on OBIS version of Maritimes RV data
- Thesis chapter on demographic consequences of changes in growth and maturation (using US and Maritimes survey data)
- Thesis chapter on density-dependent habitat selection models (using US and Maritimes survey data)

Maritimes Atlas developed during 2012 casual contract, published in 2013

- Published as a Technical Report
- SQL queries to obtain SUMMER survey information directly from Oracle
- R routines to analyse data and produce figures
- make, Perl and LaTeX to assemble the Technical Report
- start the script before you go home, come back the next morning to a fresh document, fix and improve things, repeat

Hugo Bourdages from IML shared his code (used for his Atlas) with me and I used it as a starting point to develop the analytical framework required to produce the Atlas

2020 collaboration with Maritimes to update Atlas

- I am currently working on a similar document for the annual Gulf RV survey
- A bit of self-promotion on my part perked Catalina's interest in the 2013 Atlas
- A few conversations later and we had a collaborative platform established to update the 2013 Atlas
- Improved version of 2013 framework, based on what I learned since, learned from others (e.g. Sean Anderson), and from recent software development
- Draft document ready

Tools

- Moving beyond spreadsheets, use R
- Moving beyond GIS graphical user interface, use R
  - Example: analysis of lobster fishing effort in the Gulf of St. Lawrence
- Moving beyond sending email attachments and using network drives, use git
- Moving beyond Microsoft Word to produce documents, use csasdown

Recommendations / Lessons learned

- Mapping products should be the starting point, not the end point
- Nothing will ever replace sound statistical approaches
- Reproducibility is not achievable if it involves ad-hoc clicking and sequential steps
- while I love doing it, it is not necessary to build everything from the ground up every time, hence the separation of data extractions and generation of figures in the current Atlas
- git (and its associated helpers like GitHub and GitLab) is the platform of choice for collaboration
  - yes, it is intimidating
  - but, knowing 5 commands (status, add, commit, push, pull) is enough to be functional

Next steps

Peer-review of Atlas methods and outputs - What is/are the deliverable(s), under a CSAS process?

Integration of Mike McMahon's ecosystem of R packages for the Maritimes in the Atlas

Multi-species / functional groups analyses (e.g. fish diversity, trends in groups, . . .)

The methodology developed for the Maritimes Atlas could be expanded to include:

- Georges Bank winter survey, deep strata in summer survey, historical surveys, . . . other surveys of interest
- other monitoring programs
- . . . any yearly/routine data collection program can benefit from this repeatable/reproducible approach