

ag comments

Generelt

Flott arbeid! Kjekt å se alt dere har lært. Dere bruker verktøyene som er tilgjengelige i Quarto slik som kryssreferanser og siteringssystemet. Mye fin kode, men jeg vil anbefale dere å bruke flere chunker. En chunk for hver kommando/pipe. Dette gir større fleksibilitet mht. hva dere vil vise i det endelige dokumentet. Setter dere label: på chunkene er det også mye lettere å finne eventuelle feil.

Diskusjonen og analysen syntes jeg var meget bra.

some points

- Jeg anbefaler dere sterkt å benytte quarto sitt innebygde system med overskrifter (seks ulike nivå). Da kan dere kryssreferere til ulike avnitt, slå av og på nummerering i YAML, automatisk generere innholdsfortegnelse etc
- Satt lang: nb så får vi norsk ordeling i pdf format og også norske ord for Figur, Tabell etc.
- Dere leser inn `kc_house_data` og setter riktig projeksjon helt perfekt. Jeg har satt argumentet `show_col_types = FALSE` så dere slipper gnål om vektor typer i endelig dokument.
- Beregningen `dist_cdb` er veldig elegant.
- Satt argumentet `quiet = TRUE` på `st_read()` for innlesing av WADOH kartet slik at meldingene derfra ikke forurensar det endelige dokumentet.
- Dere behandler også WADOH kartet perfekt.
- Jeg anbefaler dere å benytte en chunk per kommando/pipe. Setter dere også label på chunkene blir det mye lettere å finne eventuelle feil.
- Se bruken av `starts_with()` under innlesing av income data. Disse hjelpefunksjonene fra `tidyselect` pakken kan spare masse skriving.
- Å legge `race` og `income` til `kc_wadoh_map` gjør dere unødvendig tungvindt. Kan gjøre direkte vha. `to_left_join()`
- Perfekt bruk av `spatial join` for å overføre område data til husdataene.
- Fin bruk av `na.omit` for å fjerne obeservasjoner med NA verdier.

- `tracts10_join[complete.cases(tracts10_join),]` er en annen måte for å oppnå det samme, men `na.omit()` er jo besnærende kortfattet. Fordelen med `complete.cases()` er at man kan angi hvilke kolonner man vil sjekke for NA verdier.
- Satt `quiet = TRUE` for `st_write(tracts10_join, "house_data.gpkg" ...` siden vi ikke trenger denne infoen i det endelige dokumentet
- Det går an å kryssreferer til figurer som er satt inn med `img` tag. Se hva jeg har gjort for fig-2. Dere legger inn `caption` i dialogboksen og så må dere under fanen `Attributes` legg inn fig-2 som ID.
- Ser at dere har brukt bivariante LISA, tror jeg ville også ha tatt med plain (univariate) LISA med salgspris som variabel.
- Under Hedoniske modeller: Lange uttrykk er noe herk. Se forslag til løsning vha. `\begin{aligned} \end{aligned}`
- Vil også foreslå å bruke LaTeX og tilhørende `()` for å nummerere uttrykkene. Skal virke med pdf og html output format. Se [her](#) for detaljer.
- Satte vidden på første kolonne i den store regresjonstabellen vha. `width(j = 1, width = 2)`
- Satte også en `autofit()`
- Satte `ft.arraystretch: 1.2` så blir ikke tabellen fullt så kompakt
- Plot av regresjonsmodeller vha. `plot(lm(mod1))`. Vær klar over which argumentet slik at hvis dere bare trenger fitted values mot residual kan dere skrive `plot(lm(mod1), which = 1)`
- for `table_data1` satte jeg antall desimaler til 3. Det er mer enn nok. Tilsvarende for `table_data2` og `table_data3`.
- Anselin: Her ser vi et problem med Anselins metode Rao's score gir ikke noe klart svar på hvilken modell vi bør velge.
- Siden dere konkluderer med at SAR er den beste ville jeg bare droppet estimering av SAR modellen
- Dropp summary av SAR modellen. For denne er det impacts som skal tolkes.
- Det som er av interesse i summary er estimatet av rho. Den kan vi plukke ut fra summary vha. en egenskap ved `map()` funksjonen. Lar vi `.f` argumentet være en tekststreng vil `map()` hente denne frem fra en liste vha. `pluck()` funksjonen.
- For SAR, SDEM og SLX er det impacts som skal tolkes. Det beste er å ikke gjengi summary. Eventuelt bare plukke ut estimert autokorrelasjonsparameter fra summary.
- For SEM og OLS kan man bruke `summary()` og tolke regresjonskoeffisienter på vanlig måte.
- LeSage: Hvis det er snakk om et lokalt fenomen så er det SDEM, SEM, SLX og OLS som er de relevante modellen (SAR er ikke en begrenset versjon og har ikke noe her å gjøre). SEM, SLX og OLS er begrensede («restricted») versjoner av SDEM. OLS er begrenset versjon av SEM og også av SLX. Siden dette er restricted versjoner kan de testes mot hverandre vha. likelihood ratio test og i de fleste tilfeller vil vi få et klart svar på hvilken modell vi bør velge. Unntaket er hvis vi finner at både SEM og SLX er bedre enn SDEM og at både SEM og SLX er bedre enn OLS. Da har vi ikke noe klart for hvilken av SEM og SLX som er best (siden disse ikke er restricted versjoner av hverandre). Da må vi ty til likelihood-verdier, AIC eller liknende kriterier.