CODECHECK certificate 2022-001

https://doi.org/10.5281/zenodo.FIXME

Item	Value
Title	Geographically Weighted Regressions for prioritizing educational planning, policies, and interventions
Authors	Amélie Gagnon, Germán Vargas Mesa
Reference	IIEP technical note (2021) http://www.iiep.unesco.org/en/publication/geographically-weighted-regressions-prioritizing-educational-planning-policies-and
Codechecker	Stephen J. Eglen
Date of check	2022-01-19 10:00:00
Summary	R code for this project was reproducible.
Repository	https://github.com/codecheckers/GWR-in-educational-planning

Table 1: CODECHECK summary

```
## Warning in sprintf("\\href{%s}{\\path{%s}}", urls, m[,
## 1], m[, 1]): one argument not used by format '\href{%s}
## {\path{%s}}'
```

Output	Comment	Size (b)
codecheck/results/fig4.pdf	manuscript Figure 4	33975
codecheck/results/fig8.pdf	manuscript Figure 8	11040
codecheck/results/table3.csv	Data underlying manuscript Table 3	1445
codecheck/results/code.Rout	Text output from R code (not in manuscript)	54601
codecheck/results/Rplots.pdf	Graphical output from R code (not in manuscript)	75407993

Table 2: Summary of output files generated

Summary

The code could re-run.

Key challenges: (1) dataset required is large and access is provided on request rather than being freely available on the internet. (2) significant number of R packages to install – but all are available and so just requires some time to set up, together with corresponding unix binaries.

Questions

1. Key figures from the manuscript (figures 5–7) are not directly output from R, so could not be regenerated directly. Is there an easy way for them to be regenerated (even just the core data, not necessarily with all the extra markup and context)? I have included Rplots.pdf in case any of those plots relate to the plots in the paper. (Click on the name in the table to see the file online.)
2. Do either of you have orcids?
3. See some minor problems below with Table 3.
4. Could Table 1 easily be regenerated? I couldn't see it being generated by the R code.

CODECHECKER notes

The GitHub repo was https://github.com/codecheckers/GWR-in-educational-planning

Installation prerequisites

A file `codecheck/installs.R` was created to do the local installations. Some of the R packages required extra linux packages to be installed, notably 'gdal' and 'udunits' – see the R script for details. The installation required many R packages, taking about 20 minutes to install. This was non-trivial to setup, and perhaps in future could benefit from a Docker container.

Data

The data required for the project is not publicly available, but is available upon request (note procedure in github file). The file `Replication files.zip` is 867 Mb. This zip file was unpacked and stored in a separate directory to the github.

As the main R script was quite long, I created a symlink to the file

```
ln -s "Geographically weighted regressions for prioritizing educational planning, policies, and interver
```

Inside `code.R`, I set the variable

```
replication.folder = "/home/stephen/archive/proj/2022/gwr/Replication files"
```

Changes to the code.

Only minor changes to the code were required. I used `file.path()` rather than assuming the directory separator is \\ (as it is on Windows). I also added the followin line to the end of the R script so that it reports the packages used in the R at the end of code.Rout:

```
sessionInfo()
```

Running the code

To run the code:

```
R CMD BATCH code.R
```

I then used a script, `running.sh`, to copy key outputs across from the directory where code was run into the codecheck repo. The code took just over an hour to run. I have stored the `Rplots.pdf` and `code.Rout` output from the run into the results directory.

View of GWR model selection with different variables

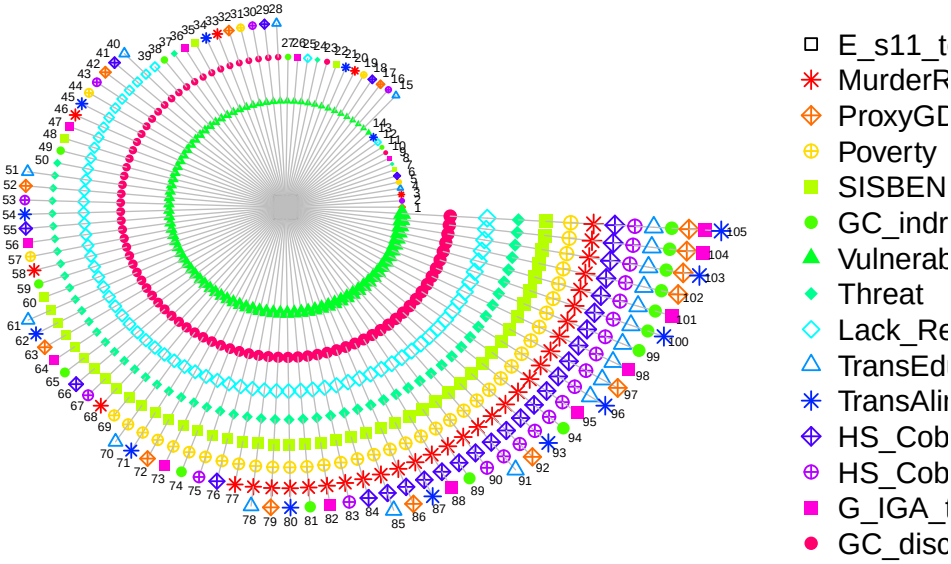

Figure C1: manuscript Figure 4

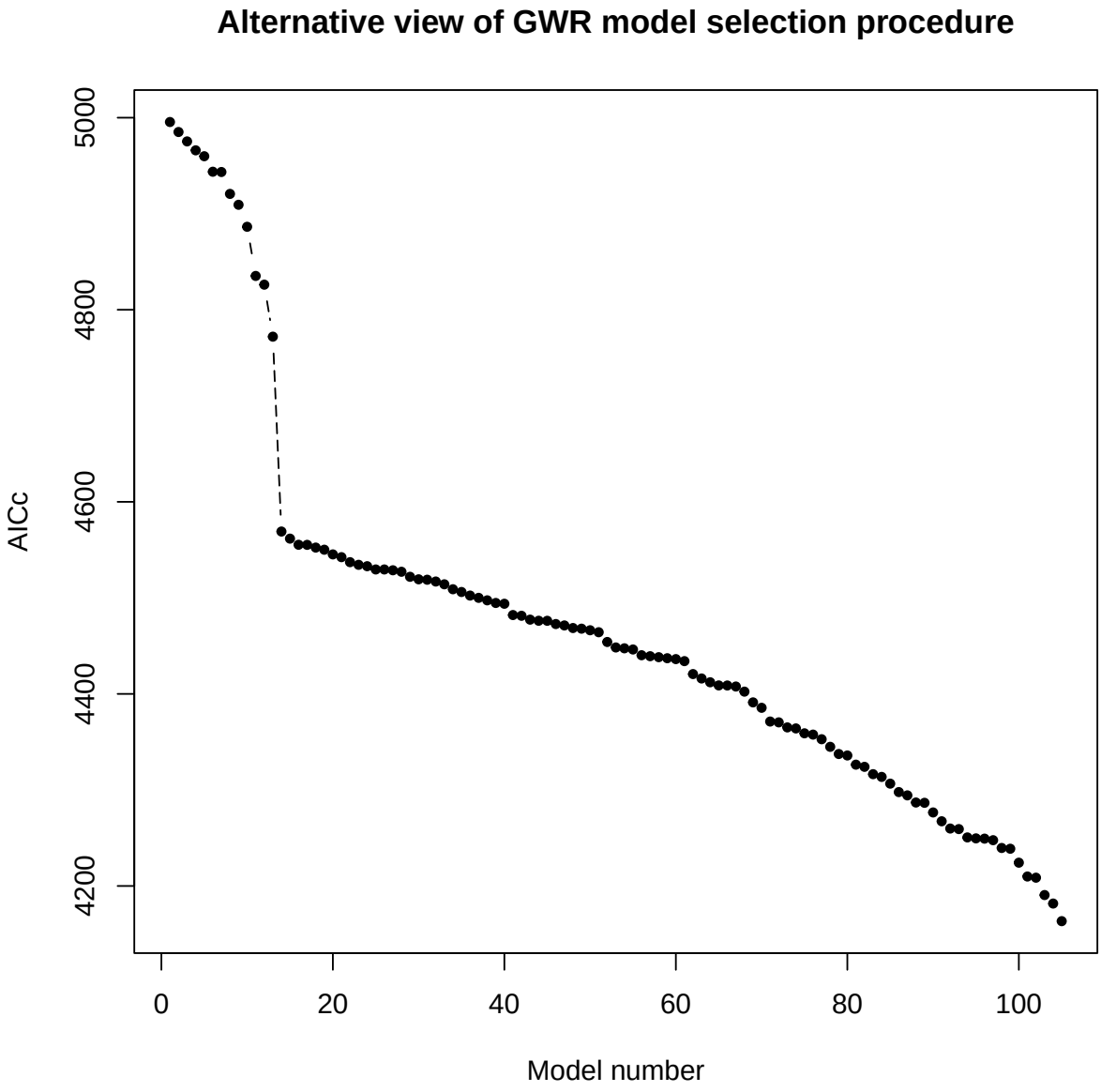

Figure C2: manuscript Figure 8

Table 3

Table 3 was reproducible. The .csv file underlying the outputs was saved into the `Replication files/Tables/` directory. This can be read into R and rendered.

```
require(xtable)
tab3 <- read.csv("results/table3.csv")
stars <- rep("", nrow(tab3))
p <- tab3$p.value

## order important here -- do least significant first.
if (any(sig <- (p < 0.1) ))   stars[sig] <- "*"
if (any(sig <- (p < 0.05) ))   stars[sig] <- "**"
if (any(sig <- (p < 0.01) ))   stars[sig] <- "***"
tab3$significance <- stars

## rearrange the rows to match the paper
reorder <- c(2, 14, 15, 12, 5, 7, 6, 8, 4, 13, 9, 10, 11, 3, 1)

tab3_neat <- tab3[reorder, c(2, 3, 4, 7)]
```

term	estimate	std.error	significance
Vulnerabil	-6.474E-01	8.082E-02	***
G_IGA_tota	6.908E-02	9.178E-03	***
TransAlimE	-2.082E-04	4.513E-05	***
GC_indrura	1.367E+00	4.869E-01	***
Threat	1.207E-01	5.724E-02	**
Poverty	4.197E-02	7.559E-03	***
SISBEN1PC	-1.773E+01	3.871E+00	***
MurderRate	-1.156E-02	2.808E-03	***
Lack_Respo	-5.360E-02	7.173E-02	
ProxyGDP	3.229E-01	1.293E-01	**
HS_Cober_7	5.764E-03	3.870E-03	
HS_Cober_1	-5.864E-03	3.409E-03	*
TransEducP	5.273E-07	1.091E-06	
GC_discapi	2.757E-04	1.494E-03	
(Intercept)	4.564E+01	9.593E-01	***

Table C1: Reproduction of Table 3.

The R-squared value at the bottom of Table 3 in the manuscript is confirmed in the .Rout file (line 541). The number of observations can be derived from 1055 d.f. with 14 variables.

There are however two minor problems with the table:

1. *Threat* has one star in the manuscript, yet it should have two stars according to the legend.
2. *TransEducP* in the manuscript is missing the exponent; the values in Table C1 match the output from the .Rout (5.273e-07, 1.091e-06) line 532.

Acknowledgements

Citing this document

Stephen J. Eglen (2022). CODECHECK Certificate 2022-001. Zenodo. https://doi.org/10.5281/zenodo.FIXME

About CODECHECK

This certificate confirms that the codechecker could independently reproduce the results of a computational analysis given the data and code from a third party. A CODECHECK does not check whether the original computation analysis is correct. However, as all materials required for the reproduction are freely available by following the links in this document, the reader can then study for themselves the code and data.

About this document

This document was created using R Markdown using the `codecheck` R package. `make codecheck.pdf` will regenerate the report file.

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Manjaro Linux
##
## Matrix products: default
## BLAS:   /usr/lib/libopenblasp-r0.3.19.so
## LAPACK: /usr/lib/liblapack.so.3.10.0
##
## locale:
##  [1] LC_CTYPE=en_GB.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_GB.UTF-8        LC_COLLATE=en_GB.UTF-8
##  [5] LC_MONETARY=en_GB.UTF-8    LC_MESSAGES=en_GB.UTF-8
##  [7] LC_PAPER=en_GB.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets
## [6] methods   base
##
## other attached packages:
##  [1] readr_2.1.1          tibble_3.1.6
##  [3] xtable_1.8-4         yaml_2.2.1
##  [5] rprojroot_2.0.2      knitr_1.37
##  [7] codecheck_0.0.0.9005 parsedate_1.2.1
##  [9] R.cache_0.15.0       gh_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] magrittr_2.0.1    hms_1.1.1         R6_2.5.1
##  [4] rlang_0.4.12      fastmap_1.1.0     fansi_0.5.0
##  [7] highr_0.9         stringr_1.4.0     httr_1.4.2
## [10] tools_4.1.2       xfun_0.29         R.oo_1.24.0
```

```
## [13] utf8_1.2.2        cli_3.1.0         htmltools_0.5.2
## [16] ellipsis_0.3.2    digest_0.6.29     lifecycle_1.0.1
## [19] crayon_1.4.2      tzdb_0.2.0        vctrs_0.3.8
## [22] R.utils_2.11.0    evaluate_0.14     rmarkdown_2.11
## [25] stringi_1.7.6     pillar_1.6.4      compiler_4.1.2
## [28] R.methodsS3_1.8.1 jsonlite_1.7.2    pkgconfig_2.0.3
```