

CODECHECK certificate 2020-016

<https://doi.org/10.5281/zenodo.3981253>






Item	Value
Title	Opening practice: supporting reproducibility and critical spatial data science
Authors	Chris Brunsdon  , Alexis Comber 
Reference	https://doi.org/10.1007/s10109-020-00334-2
Codechecker	Daniel Nüst 
Date of check	2020-06-02
Summary	A small R script to render a map and two tables. Minor code adjustments were made, but reproduction of results (one figure, two tables) was successful.
Repository	https://github.com/codecheckers/OpeningPractice

Table 1: CODECHECK summary

output	comment	size
figure1.png	Figure 1: Housing data and different census areas scales ...	460634
table2.md	Table 2: The model coefficient estimates for the individual input variables ...	407
table3.md	Table 3: The variable importance (expressed as a percentage) ...	322

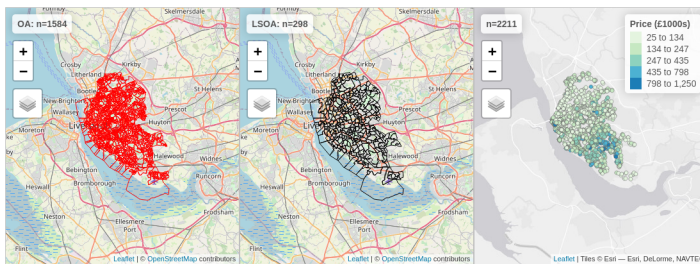
Table 2: Summary of output files generated

Summary

I could reproduce one figure and two tables from the paper. The code required some small fixes, such as a missing `library()` statement. I also had to manually create a screenshot of Figure 1, but based on a visual inspection the figures from the paper match the ones in the repository and the ones recreated by me. The numbers in reproduced Tables 2 and 3 match the ones in the paper with only small negligible numerical differences on some values.

Reproduction of Figure 1

Screenshot of interactive output.



Reproduction of Table 2

```
readLines("table2.md")
```

```
## [1] ""
## [2] ""
## [3] "|          |Covariate |      OA|      LSOA|"
## [4] "|:-----|:-----|-----:|-----:|"
## [5] "|(Intercept)|(Intercept)| 31.653| -43.505|"
## [6] "|gs_area   |gs_area   |  0.873|  0.412|"
## [7] "|u25       |u25       |  1.994|  2.882|"
## [8] "|u45       |u45       |  0.739|  1.962|"
## [9] "|u65       |u65       |  5.388|  5.543|"
## [10] "|o65       |o65       |  3.496|  6.967|"
## [11] "|unplyd    |unplyd    | -8.168| -10.850|"
```

Reproduction of Table 3

```
readLines("table3.md")
```

```
## [1] ""
## [2] ""
## [3] "|          |Covariate |      OA|      LSOA|"
## [4] "|:-----|:-----|-----:|-----:|"
## [5] "|gs_area |gs_area | 25.368|  0.000|"
## [6] "|u25     |u25     | 17.350| 12.935|"
## [7] "|u45     |u45     |  0.000|  3.070|"
## [8] "|u65     |u65     | 37.452| 15.192|"
## [9] "|o65     |o65     | 31.844| 35.560|"
## [10] "|unplyd |unplyd | 100.000| 100.000|"
```

CODECHECKER notes

Since the authors were not aware of CODECHECK at the time of submission, I did the following preparation steps:

- source the data file locally, because I have no control over the lexcomber/OpeningPractice repository
- manually set the bbox of the plots and enable map sync, so the views match each other; the default for the third plot mismatches the polygon data, because `st_bbox(props_oa)` includes (0,0) as a corner; adjusted the zoom level to more closely match the paper's figure
- saved Figure 1 from the PDF to a file, so it can be added to the manifest
- saved Tables to files so they can be added to the manifest, manually transferring the values from the paper

For details of the preparation steps see commit `7f52eb2b99087fedd5db0d72f7cea32ddc610013`.

Then I continued with the actual CODECHECK.

When starting the check, I had problems installing all required libraries locally (Ubuntu 19.10), where `rgdal` could not be updated when I wanted to install `tmap`. Therefore I switched to an `rocker/geospatial` container with R 4.0.0, which I started with the following command:

```
docker run --rm -it -p 8787:8787 -e PASSWORD=simple \  
-v $(shell pwd):/home/rstudio/OpeningPractice rocker/geospatial:4.0.0
```

From the required libraries, only the `repmis` package was missing, so I added it to a file `codecheck/install.R`. Line 31 gave me the following error:

```
TopologyException: Input geom 1 is invalid: Ring Self-intersection at or near point
```

Based on this issue, I wrapped `oa` in `sf::st_make_valid(oa)`, and the error goes away. This was not needed for `lsoa`. I made this change directly in `github_script.R`. I continued to execute commands line by line, until the function `train()`, which was not available. I needed to install and load the package `caret`. I made this change directly in `github_script.R`.

With these changes, I could `source()` the whole script file and saved the generated `tableX.md` files into the `codecheck` directory. The whole script only takes a few moments to run on my computer.

General feedback on the code

- I suggest to make the map titles dependent on the data, i.e. not hardcoding “n=1584” but using `nrow(oa)`.
- The code would also be more readable with more consistent formatting and a few new lines.
- The data should be saved in a more accessible file format, not as a binary `.RData` file; a quick test saving as GeoJSON resulted in a marginally larger but plain text file not limited to R users.
- The maps should be saved from the code, not a screenshot of the interactive view.

Colophon

How to cite this report

Daniel Nüst. (2020, June 2). CODECHECK certificate 2020-016. Zenodo. <http://doi.org/10.5281/zenodo.3873153>

Environment used to create the report (not conduct the check)

```
sessionInfo()
```

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/liblapack.so.3
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
## [3] LC_TIME=de_DE.UTF-8 LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=de_DE.UTF-8 LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=de_DE.UTF-8 LC_NAME=C
## [9] LC_ADDRESS=C LC_TELEPHONE=C
## [11] LC_MEASUREMENT=de_DE.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats graphics grDevices utils datasets
## [6] methods base
##
## other attached packages:
## [1] readr_1.3.1 tibble_3.0.3
## [3] xtable_1.8-4 yaml_2.2.1
## [5] rprojroot_1.3-2 knitr_1.29
## [7] codecheck_0.0.0.9005 parsedate_1.2.0
## [9] R.cache_0.14.0 gh_1.1.0
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.5 magrittr_1.5 hms_0.5.3
## [4] R6_2.4.1 rlang_0.4.7 fansi_0.4.1
## [7] stringr_1.4.0 httr_1.4.2 tools_4.0.2
## [10] xfun_0.15 R.oo_1.23.0 cli_2.0.2
## [13] ellipsis_0.3.1 htmltools_0.5.0 assertthat_0.2.1
## [16] digest_0.6.25 lifecycle_0.2.0 crayon_1.3.4
## [19] vctrs_0.3.2 R.utils_2.9.2 glue_1.4.1
## [22] evaluate_0.14 rmarkdown_2.3 stringi_1.4.6
## [25] pillar_1.4.6 compiler_4.0.2 backports_1.1.8
## [28] R.methodsS3_1.8.0 jsonlite_1.7.0 pkgconfig_2.0.3
```