Lab 1

Antonio Jurlina

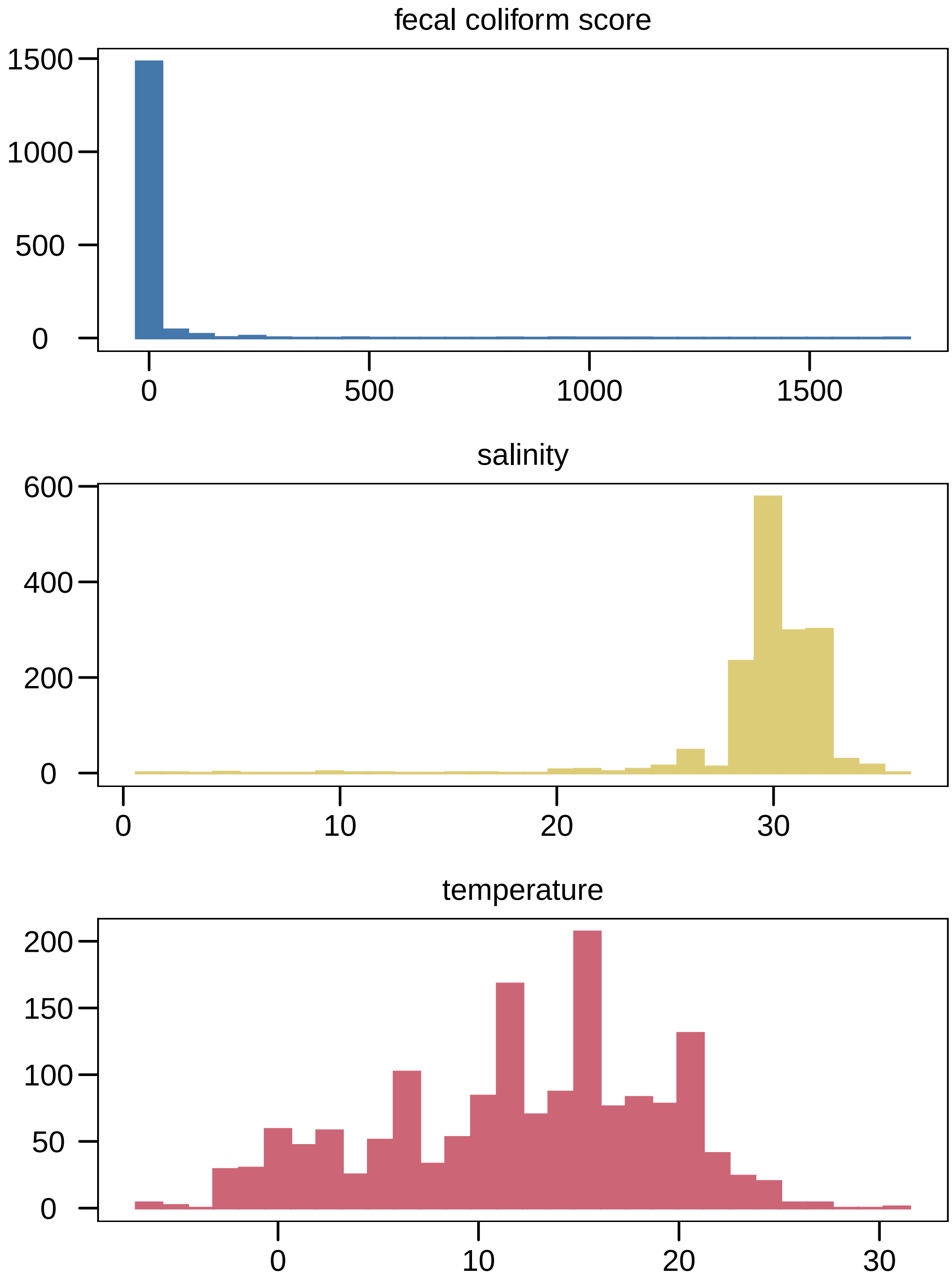
09/08/2020

Data Summary

variable	min	average	max	variance	kurtosis	skewness
salinity	1.5	29.95818	36.0	6.911182	37.1530610	-4.6868485
temperature	-7.0	12.12878	30.3	47.175415	-0.6340376	-0.3904822

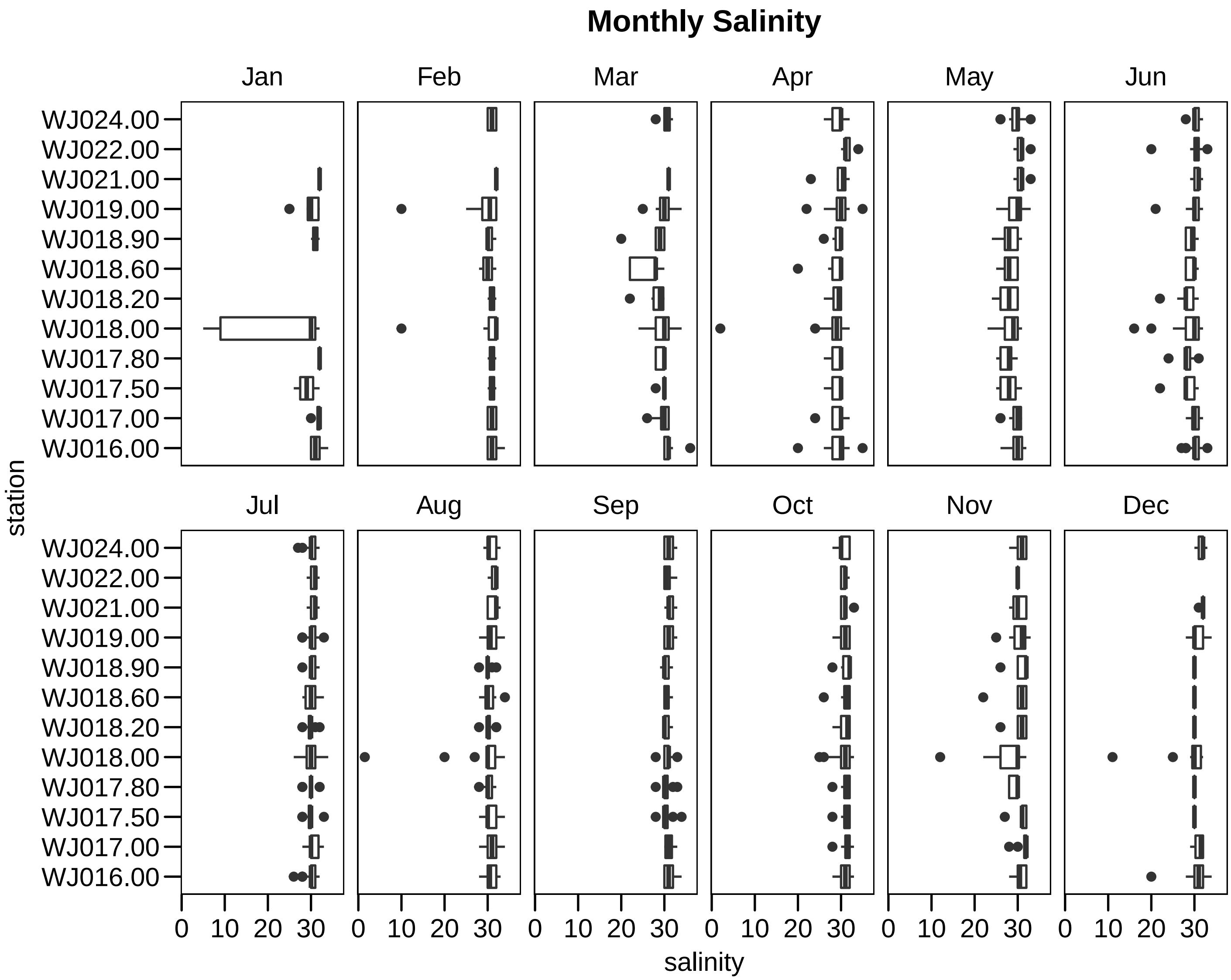
Looking at the summary data we can expect the distribution of monthly temperatures (over many years and measuring stations) to be fairly symmetrical due to the low skewness and kurtosis. These values would imply higher spread around the average, as evident by the reported variance. For salinity, extremely high kurtosis indicates much data falling on, or near, the average, further confirmed by low variance. This distribution is skeewed left. Given the similar number of observations range they encompass, comparing them to one another holds up in this particular context. Most salinity variability is expected to be location based. Measurements nearer to freshwater sources are going to have constantly lower readings. Temperature we expect to vary seasonally and also much more than salinity. This fact is reflected in the data summary.

Histogram



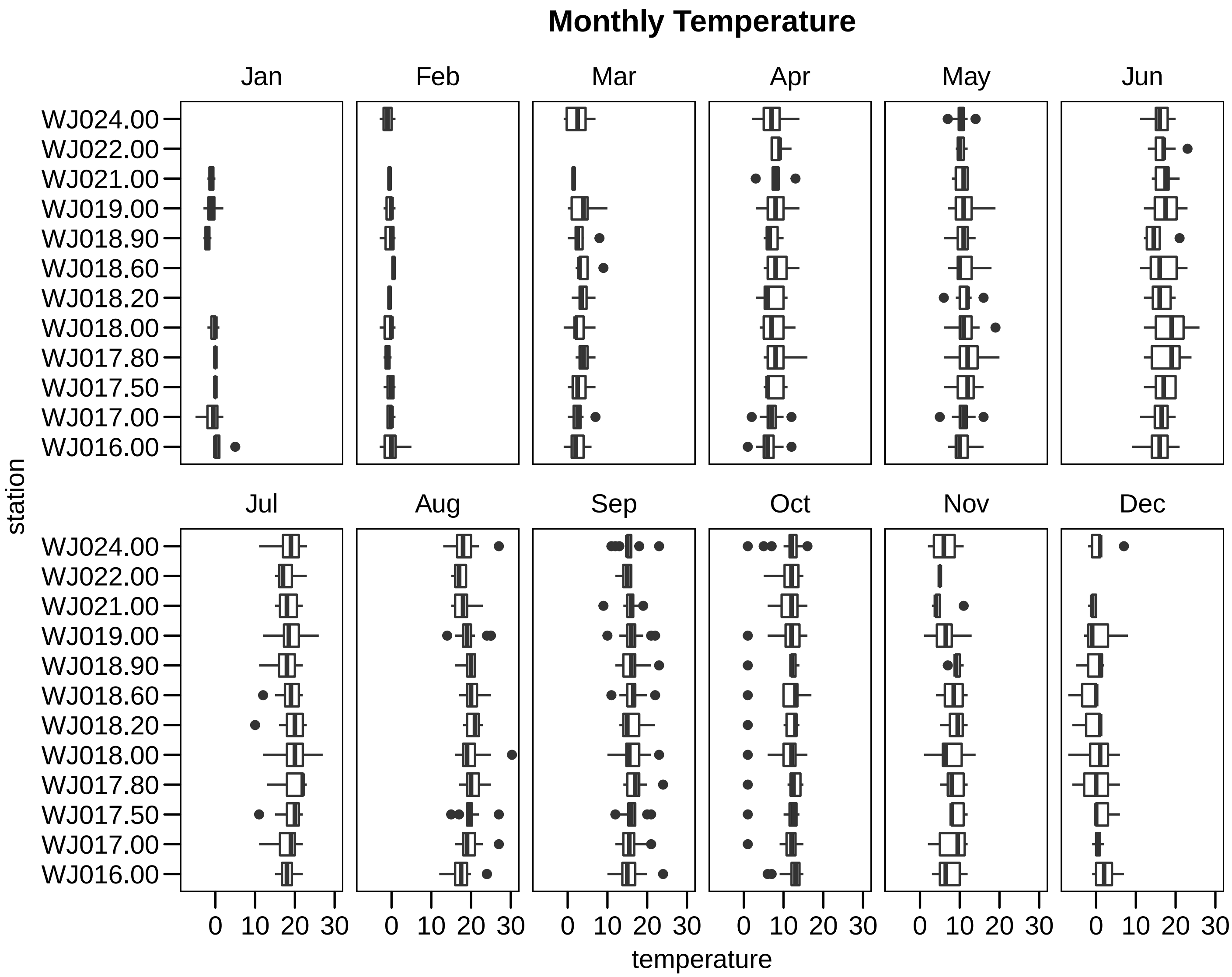
Fecal coliform score distribution is heavily skeewed right, a good indicator of water cleanliness as most values are indeed grouped around 0, with several severe outliers. Given the detail of the data provided, these should be easy to pinpoint. Salinity measurements form a distribution that skews slightly left, with a sharp peak around the average (high kurtosis) as there is not much variance to this data (except certain outliers). The pattern here, as noted before, is likely to be spatial. Measurements performed near locations where mixing of fresh water with the sea is occurring are bound to have much lower salinity scores. Finally, the distribution of temperature measurements is distributed fairly symmetrically with higher variance that the previous two. This variance can be attributed to temporal changes, as tempertures measured across different months will reflect the expected seasonal patterns.

Salinity Boxplot



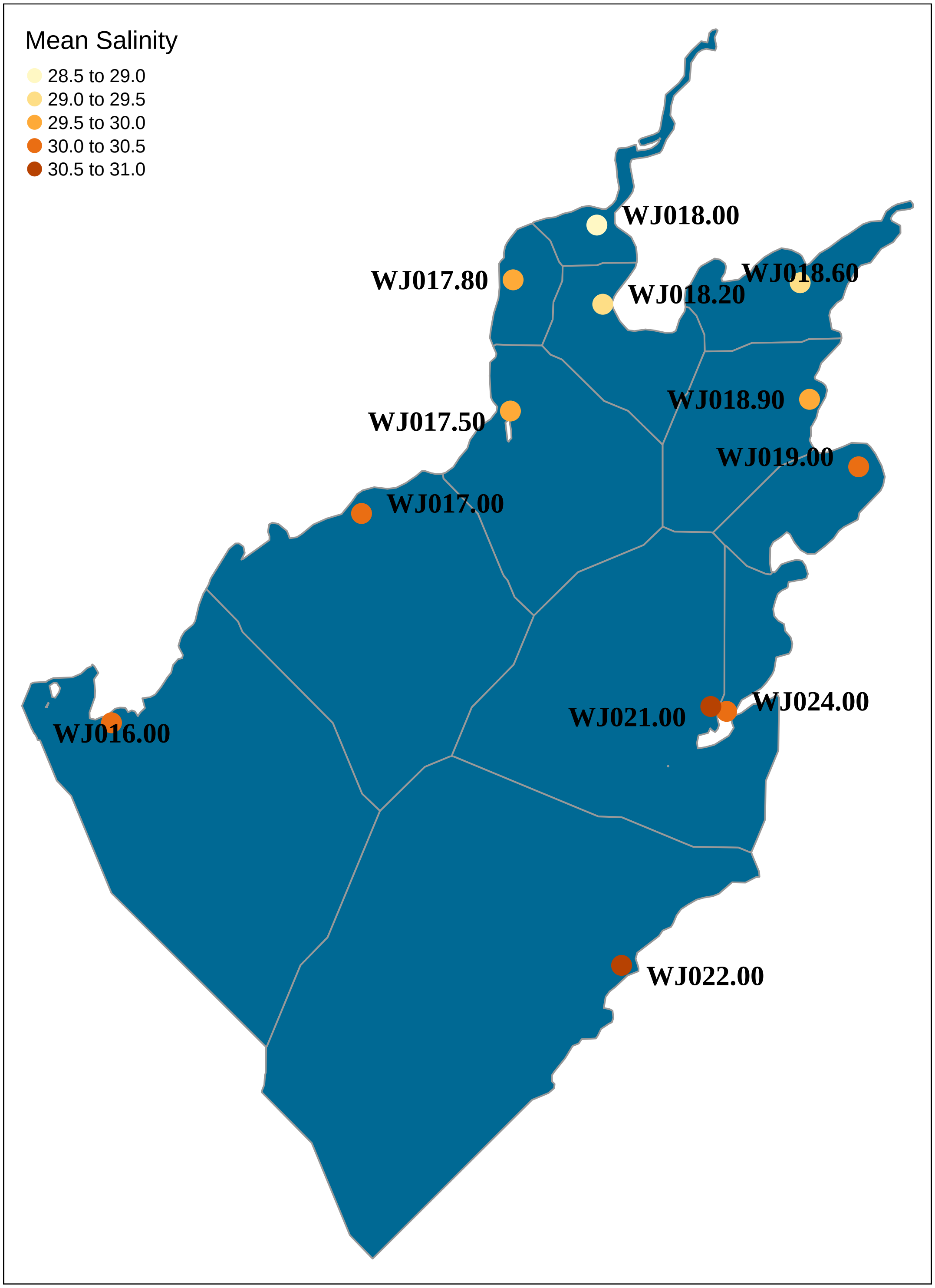
The salinity boxplots give further evidence to claims brought before - this data is highly centralized around a stable mean, regardless of the season, varying mostly by station. One of the stations (WJ018.00) reports a fairly unusual streak of outliers but this is to be expected as well, given it's location. The map (shown below) confirms this station as positioned quite close to a river mouth. This would account for the low salinity scores.

Temperature Boxplot



The temperature boxplots depict a clear temporal pattern. As time flows from January onward to August, average temperatures keep rising. After this point, they slowly return back to where they started, as winter settles back in. There are outliers, but all stations move around in unison, on average

Map



Mean Salinity
28.5 to 29.0
29.0 to 29.5
29.5 to 30.0
30.0 to 30.5
30.5 to 31.0
WJ018.00
WJ017.80
WJ018.20
WJ018.60
WJ018.90
WJ017.50
WJ019.00
WJ017.00
WJ021.00
WJ024.00
WJ016.00
WJ022.00

```r
sessionInfo()
```

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] e1071_1.7-3     rgdal_1.5-16    rebus_0.1-3     sp_1.4-2
##  [5] sf_0.9-5        tmap_3.1        here_0.1        ggthemes_4.2.0
##  [9] lubridate_1.7.9 forcats_0.5.0   stringr_1.4.0   dplyr_1.0.2
## [13] purrr_0.3.4     readr_1.3.1     tidyr_1.1.2     tibble_3.0.3
## [17] ggplot2_3.3.2   tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] fs_1.5.0              RColorBrewer_1.1-2    httr_1.4.2
##  [4] rprojroot_1.3-2       tools_3.6.2           backports_1.1.9
##  [7] R6_2.4.1              KernSmooth_2.23-17    DBI_1.1.0
## [10] rebus.base_0.0-3      colorspace_1.4-1      raster_3.3-13
## [13] withr_2.2.0           tidyselect_1.1.0      leaflet_2.0.3
## [16] compiler_3.6.2        leafem_0.1.3          cli_2.0.2
## [19] rvest_0.3.6           xml2_1.3.2            labeling_0.3
## [22] scales_1.1.1          classInt_0.4-3        digest_0.6.25
## [25] rebus.unicode_0.0-2   rmarkdown_2.3         base64enc_0.1-3
## [28] dichromat_2.0-0       pkgconfig_2.0.3       htmltools_0.5.0
## [31] highr_0.8             dbplyr_1.4.4          htmlwidgets_1.5.1
## [34] rlang_0.4.7           readxl_1.3.1          rstudioapi_0.11
## [37] farver_2.0.3          generics_0.0.2        jsonlite_1.7.0
## [40] crosstalk_1.1.0.1     magrittr_1.5          Rcpp_1.0.5
## [43] munsell_0.5.0         fansi_0.4.1           abind_1.4-5
## [46] lifecycle_0.2.0       stringi_1.4.6         leafsync_0.1.0
## [49] yaml_2.2.1            tmaptools_3.1         grid_3.6.2
## [52] blob_1.2.1            parallel_3.6.2        crayon_1.3.4
## [55] rebus.datetimes_0.0-1 lattice_0.20-41       stars_0.4-3
## [58] haven_2.3.1           hms_0.5.3             knitr_1.29
## [61] pillar_1.4.6          rebus.numbers_0.0-1   codetools_0.2-16
## [64] reprex_0.3.0          XML_3.99-0.3          glue_1.4.2
## [67] evaluate_0.14         modelr_0.1.8          vctrs_0.3.4
## [70] png_0.1-7             cellranger_1.1.0      gtable_0.3.0
## [73] assertthat_0.2.1      xfun_0.16             lwgeom_0.2-5
## [76] broom_0.7.0           class_7.3-17          viridisLite_0.3.0
## [79] units_0.6-7           ellipsis_0.3.1
```