wqpcleanr Vignette

Kevin Hitt

2020-04-25

```r
library(wqpcleanr)
library(ggplot2)
library(ggmap)
#> Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
#> Please cite ggmap if you use it! See citation("ggmap") for details.
```

```r
# Load data from WQP portal -----------------------------------------------
#
# Please provide 1 start date and 1 end date for the search parameters.
#
# The data import is handled automatically by the package via direct URL download.
#
# Column names are automatically assigned as:
#                   org_id
#                   org_name
#                   monitor_id
#                   monitor_name
#                   monitor_type
#                   monitor_descrip
#                   light_code
#                   drainage_area
#                   drainage_area_unit
#                   cont_drainage_area
#                   cont_drainage_unit
#                   lat
#                   lon
#                   map_scale
#                   horiz_acc
#                   horiz_acc_unit
#                   horiz_source
#                   horiz_datum
#                   vert_measure
#                   vert_measure_unit
#                   vert_acc
#                   vert_acc_unit
#                   vert_source
#                   vert_datum
#                   country_id
#                   state_id
#                   county_id
#                   aquifer
#                   aquifer_formation
```

```r
#                  aquifer_type
#                  construction_date
#                  well_depth
#                  well_depth_unit
#                  wellhole_depth
#                  wellhole_depth_unit
#                  provider
#
#
# Currently, data is specific to Hillsborough County, Florida
wqp <- load_wqp("12-31-2010","12-31-2020")

# Inspect loaded data
str(wqp)
#> 'data.frame':    44 obs. of  36 variables:
#>  $ org_id             : Factor w/ 1 level "USGS-FL": 1 1 1 1 1 1 1 1 1 1 ...
#>  $ org_name           : Factor w/ 1 level "USGS Florida Water Science Center": 1 1 1 1 1 1 1 1 1 1 .
#>  $ monitor_id         : Factor w/ 44 levels "USGS-02300100",..: 1 2 3 4 5 6 7 8 9 10 ...
#>  $ monitor_name       : Factor w/ 44 levels "ALAFIA RIVER AT BELL SHOALS NEAR RIVERVIEW FL",..: 23 6
#>  $ monitor_type       : Factor w/ 4 levels "Estuary","Spring",..: 3 3 3 3 3 3 3 3 1 3 ...
#>  $ monitor_descrip    : logi  NA NA NA NA NA NA ...
#>  $ light_code         : int  3100203 3100206 3100204 3100204 3100204 3100204 3100204 3100204 3100206
#>  $ drainage_area      : num  31.4 29.1 135 107 376 414 419 418 36.8 110 ...
#>  $ drainage_area_unit : Factor w/ 2 levels "","sq mi": 2 2 2 2 2 2 2 2 2 2 ...
#>  $ cont_drainage_area : num  31.4 29.1 NA 107 NA NA NA NA NA 110 ...
#>  $ cont_drainage_unit : Factor w/ 2 levels "","sq mi": 2 2 1 2 1 1 1 1 1 2 ...
#>  $ lat                : num  27.7 27.8 27.9 27.8 27.9 ...
#>  $ lon                : num  -82.2 -82.4 -82.1 -82.1 -82.3 ...
#>  $ map_scale          : int  24000 24000 24000 24000 24000 24000 24000 24000 24000 24000 ...
#>  $ horiz_acc          : num  5 5 5 5 1 1 1 5 1 1 ...
#>  $ horiz_acc_unit     : Factor w/ 1 level "seconds": 1 1 1 1 1 1 1 1 1 1 ...
#>  $ horiz_source       : Factor w/ 3 levels "Differentially corrected Global Positioning System.",..:
#>  $ horiz_datum        : Factor w/ 1 level "NAD83": 1 1 1 1 1 1 1 1 1 1 ...
#>  $ vert_measure       : num  44.09 -0.88 37.59 39.03 0 ...
#>  $ vert_measure_unit  : Factor w/ 2 levels "","feet": 2 2 2 2 2 2 2 2 2 2 ...
#>  $ vert_acc           : num  0.1 0.1 0.1 0.1 0.01 0.01 0.01 0.1 0.01 0.1 ...
#>  $ vert_acc_unit      : Factor w/ 2 levels "","feet": 2 2 2 2 2 2 2 2 2 2 ...
#>  $ vert_source        : Factor w/ 3 levels "","Interpolated from topographic map.",..: 3 3 3 3 3 3 3
#>  $ vert_datum         : Factor w/ 3 levels "","NAVD88","NGVD29": 2 2 2 2 3 3 2 2 3 3 ...
#>  $ country_id         : Factor w/ 1 level "US": 1 1 1 1 1 1 1 1 1 1 ...
#>  $ state_id           : int  12 12 12 12 12 12 12 12 12 12 ...
#>  $ county_id          : int  57 57 57 57 57 57 57 57 57 57 ...
#>  $ aquifer            : Factor w/ 3 levels "","Floridan aquifer system",..: 1 1 1 1 1 1 1 1 1 1 ...
#>  $ aquifer_formation  : Factor w/ 3 levels "","Floridan Aquifer System",..: 1 1 1 1 1 1 1 1 1 1 ...
#>  $ aquifer_type       : Factor w/ 4 levels "","Confined single aquifer",..: 1 1 1 1 1 1 1 1 1 1 ...
#>  $ construction_date  : int  NA NA NA NA NA NA NA NA NA NA ...
#>  $ well_depth         : num  NA NA NA NA NA NA NA NA NA NA ...
#>  $ well_depth_unit    : Factor w/ 2 levels "","ft": 1 1 1 1 1 1 1 1 1 1 ...
#>  $ wellhole_depth     : num  NA NA NA NA NA NA NA NA NA NA ...
#>  $ wellhole_depth_unit: Factor w/ 2 levels "","ft": 1 1 1 1 1 1 1 1 1 1 ...
#>  $ provider           : Factor w/ 1 level "NWIS": 1 1 1 1 1 1 1 1 1 1 ...

# Preview where NA values exist
```

```r
preview_na(wqp)
#>                    na_count percent
#> org_id                    0    0.00
#> org_name                  0    0.00
#> monitor_id                0    0.00
#> monitor_name              0    0.00
#> monitor_type              0    0.00
#> monitor_descrip          44  100.00
#> light_code                0    0.00
#> drainage_area            25   56.82
#> drainage_area_unit        0    0.00
#> cont_drainage_area       36   81.82
#> cont_drainage_unit        0    0.00
#> lat                       0    0.00
#> lon                       0    0.00
#> map_scale                 0    0.00
#> horiz_acc                 0    0.00
#> horiz_acc_unit            0    0.00
#> horiz_source              0    0.00
#> horiz_datum               0    0.00
#> vert_measure              1    2.27
#> vert_measure_unit         0    0.00
#> vert_acc                  1    2.27
#> vert_acc_unit             0    0.00
#> vert_source               0    0.00
#> vert_datum                0    0.00
#> country_id                0    0.00
#> state_id                  0    0.00
#> county_id                 0    0.00
#> aquifer                   0    0.00
#> aquifer_formation         0    0.00
#> aquifer_type              0    0.00
#> construction_date        30   68.18
#> well_depth               22   50.00
#> well_depth_unit           0    0.00
#> wellhole_depth           22   50.00
#> wellhole_depth_unit       0    0.00
#> provider                  0    0.00

# Preview unique values for each column
preview_uniques(wqp)
#> $`org_id.levels(factor(y))`
#> [1] USGS-FL
#> Levels: USGS-FL
#>
#> $`org_name.levels(factor(y))`
#> [1] USGS Florida Water Science Center
#> Levels: USGS Florida Water Science Center
#>
#> $`monitor_id.levels(factor(y))`
#>  [1] USGS-02300100      USGS-02300700      USGS-02301000
#>  [4] USGS-02301300      USGS-02301638      USGS-02301718
#>  [7] USGS-02301719      USGS-02301721      USGS-02301805
```

```
#> [10] USGS-02302500        USGS-02303000        USGS-02303330
#> [13] USGS-02303350        USGS-02303800        USGS-02304000
#> [16] USGS-02304510        USGS-02304520        USGS-02306000
#> [19] USGS-023060003       USGS-023060005       USGS-023060013
#> [22] USGS-02306028        USGS-280215082280001 USGS-280235082313502
#> [25] USGS-280305082300502 USGS-280310082291002 USGS-280341082325702
#> [28] USGS-280437082303002 USGS-280451082351102 USGS-280500082313502
#> [31] USGS-280550082312202 USGS-280605082302601 USGS-280614082325701
#> [34] USGS-280641082341901 USGS-280721082303701 USGS-280734082315402
#> [37] USGS-280746082335601 USGS-280802082275701 USGS-280823082135501
#> [40] USGS-280846082134401 USGS-280848082131001 USGS-280905082293702
#> [43] USGS-280928082315202 USGS-281016082275602
#> 44 Levels: USGS-02300100 USGS-02300700 USGS-02301000 ... USGS-281016082275602
#>
#> $`monitor_name.levels(factor(y))`
#>  [1] ALAFIA RIVER AT BELL SHOALS NEAR RIVERVIEW FL
#>  [2] ALAFIA RIVER AT GIBSONTON FL
#>  [3] ALAFIA RIVER NEAR GIBSONTON FL
#>  [4] ALAFIA RIVER NEAR US 301 AT RIVERVIEW, FL
#>  [5] BLACKWATER CREEK NEAR KNIGHTS FL
#>  [6] BULLFROG CREEK NEAR WIMAUMA FL
#>  [7] CARROLLWOOD ELEMENTARY ID-057VF099
#>  [8] CHILDREN'S CAMP 51 FT FLRD WELL NR ZEPHYRHILLS FL
#>  [9] CLAYWELL ELEM AGWQMP ID-057VF092 FL
#> [10] CORNERSTONE CHURCH 28 FT NRSD WELL NEAR ODESSA FL
#> [11] CYPRESS CREEK NEAR SULPHUR SPRINGS FL
#> [12] FAIRFIELD VILLAGE ID057VF095 FL
#> [13] FORT FOSTER FLRD WELL NEAR ZEPHYRHILLS FL
#> [14] GAITHER HIGH 38FT NRSD WELL NEAR CITRUS PARK FL
#> [15] HILLSBOROUGH R AT FOWLER AV NEAR TEMPLE TERRACE FL
#> [16] HILLSBOROUGH R AT MORRIS BR NEAR THONOTOSASSA FL
#> [17] HILLSBOROUGH R. AT I-275 BRIDGE AT SULPHUR SPGS FL
#> [18] HILLSBOROUGH RIVER AT PLATT STREET AT TAMPA FL
#> [19] HILLSBOROUGH RIVER AT ROWLETT PK DR NEAR TAMPA FL
#> [20] HILLSBOROUGH RIVER AT SULPHUR SPRINGS FL
#> [21] HILLSBOROUGH RV AT STATE PARK NR ZEPHYRHILLS, FL
#> [22] KEYSTONE PRESBYTERIAN 39FT NRSD WELL NR ODESSA FL
#> [23] LITTLE MANATEE RIVER NEAR FT. LONESOME FL
#> [24] LUTZ FIRST BAPTIST FLRD WELL NEAR LUTZ FL
#> [25] MCKITRICK SCHOOL 28 FT NRSD WELL NEAR ODESSA FL
#> [26] NORTH PRONG ALAFIA RIVER AT KEYSVILLE FL
#> [27] NW ELEMENTARY 38 FT NRSD WELL NR CITRUS PARK FL
#> [28] PALM RIVER AT MOUTH AT TAMPA FL
#> [29] ROMP 66 NEAR SULPHUR SPRINGS FL
#> [30] ROY HAYNE PARK NR SULPHUR SPGS FL FLRD
#> [31] SECTION 21 MW-UF5 FLRD WELL NEAR CITRUS PARK FL
#> [32] SICKLES HIGH SCHOOL 19' NRSD NR CITRUS PARK FL
#> [33] SOUTH PRONG ALAFIA RIVER NEAR LITHIA FL
#> [34] SULPHUR SPRINGS AT SULPHUR SPRINGS FL
#> [35] SULPHUR SPRINGS MOUTH AT SULPHUR SPRINGS FL
#> [36] SULPHUR SPRINGS RUN AT SULPHUR SPRINGS FL
#> [37] SUNLAKE BAPTIST CHURCH 14' NRSD NR LUTZ FL
#> [38] TRINITY ASSEMBLY OF GOD 21 FT NRSD WELL NR LUTZ FL
```

```
#> [39] TROUT CREEK NEAR SULPHUR SPRINGS FL
#> [40] WALKER MIDDLE SCHOOL 38' FLRD WELL NR CITRUS PK FL
#> [41] WCRWSA AGWQMP ID-057VF108 ROMP-801
#> [42] WCRWSA AGWQMP ID-057VF110 ROMP-70
#> [43] WEST VILLAGE AGWQMP ID-057VS104 FL
#> [44] WETLANDS TRAIL 43 FT FLRD WELL NEAR ZEPHYRHILLS FL
#> 44 Levels: ALAFIA RIVER AT BELL SHOALS NEAR RIVERVIEW FL ...
#>
#> $`monitor_type.levels(factor(y))`
#> [1] Estuary Spring  Stream  Well
#> Levels: Estuary Spring Stream Well
#>
#> $`monitor_descrip.levels(factor(y))`
#> factor(0)
#> Levels:
#>
#> $`light_code.levels(factor(y))`
#> [1] 3100203 3100204 3100205 3100206 3100207
#> Levels: 3100203 3100204 3100205 3100206 3100207
#>
#> $`drainage_area.levels(factor(y))`
#>  [1] 23    29.1 31.4 36.8 107  110   135   160   220   375   376   414   418   419   625
#> [16] 630   636   637   694
#> 19 Levels: 107 110 135 160 220 23 29.1 31.4 36.8 375 376 414 418 419 ... 694
#>
#> $`drainage_area_unit.levels(factor(y))`
#> [1]      sq mi
#> Levels:  sq mi
#>
#> $`cont_drainage_area.levels(factor(y))`
#> [1] 23    29.1 31.4 107   110   160   220   375
#> Levels: 107 110 160 220 23 29.1 31.4 375
#>
#> $`cont_drainage_unit.levels(factor(y))`
#> [1]      sq mi
#> Levels:  sq mi
#>
#> $`lat.levels(factor(y))`
#>  [1] 27.70475366 27.79197208 27.79669525 27.85696987 27.85891477 27.85974739
#>  [7] 27.86780308 27.8839148  27.94196627 27.9422444  28.0197412  28.0202967
#> [13] 28.02113    28.02113003 28.02113004 28.02113015 28.03640706 28.04501764
#> [19] 28.05168406 28.05446224 28.06168364 28.06251698 28.07307209 28.07473869
#> [25] 28.08125    28.08888889 28.09751556 28.0986111  28.10133333 28.10394444
#> [31] 28.11141667 28.1226111  28.12605556 28.12952778 28.1338611  28.1347222
#> [37] 28.1397222  28.1398611  28.14619444 28.1466111  28.15029044 28.15130556
#> [43] 28.1578611  28.1712222
#> 44 Levels: 27.70475366 27.79197208 27.79669525 27.85696987 ... 28.1712222
#>
#> $`lon.levels(factor(y))`
#>  [1] -82.5864722  -82.5719722  -82.5656667  -82.5491667  -82.5489854
#>  [6] -82.5317222  -82.53148498 -82.5312222  -82.5295405  -82.5225959
#> [11] -82.5102222  -82.5092622  -82.5071111  -82.5012065  -82.4937222
#> [16] -82.4931507  -82.4673168  -82.46575    -82.4654444  -82.4587057
```

```
#> [21] -82.4548166  -82.4531499  -82.45231658 -82.451761   -82.4345385
#> [26] -82.4098159  -82.4091667  -82.3842598  -82.3637032  -82.3619444
#> [31] -82.3575926  -82.3520369  -82.3198141  -82.31138889 -82.2737021
#> [36] -82.232      -82.23175318 -82.2288333  -82.2193611  -82.1978672
#> [41] -82.15       -82.1175877  -82.10008689
#> 43 Levels: -82.10008689 -82.1175877 -82.15 -82.1978672 ... -82.5864722
#>
#> $`map_scale.levels(factor(y))`
#> [1] 24000
#> Levels: 24000
#>
#> $`horiz_acc.levels(factor(y))`
#> [1] 0.1 0.5 1   5
#> Levels: 0.1 0.5 1 5
#>
#> $`horiz_acc_unit.levels(factor(y))`
#> [1] seconds
#> Levels: seconds
#>
#> $`horiz_source.levels(factor(y))`
#> [1] Differentially corrected Global Positioning System.
#> [2] Interpolated from MAP.
#> [3] Mapping grade GPS unit (handheld accuracy range 12 to 40 ft)
#> 3 Levels: Differentially corrected Global Positioning System. ...
#>
#> $`horiz_datum.levels(factor(y))`
#> [1] NAD83
#> Levels: NAD83
#>
#> $`vert_measure.levels(factor(y))`
#>  [1] -10    -1.28 -0.9  -0.89 -0.88 -0.79 -0.78 -0.76 -0.75 0      32.45 35
#> [13] 37.59 39.03 40    44.09 45    47    48    50    55    56     60    70
#> [25] 75
#> 25 Levels: -0.75 -0.76 -0.78 -0.79 -0.88 -0.89 -0.9 -1.28 -10 0 32.45 ... 75
#>
#> $`vert_measure_unit.levels(factor(y))`
#> [1]    feet
#> Levels:  feet
#>
#> $`vert_acc.levels(factor(y))`
#> [1] 0.01 0.1  2.5  5
#> Levels: 0.01 0.1 2.5 5
#>
#> $`vert_acc_unit.levels(factor(y))`
#> [1]    feet
#> Levels:  feet
#>
#> $`vert_source.levels(factor(y))`
#> [1]                           Interpolated from topographic map.
#> [3] Level or other surveyed method.
#> 3 Levels:  ... Level or other surveyed method.
#>
#> $`vert_datum.levels(factor(y))`
```

```
#> [1]       NAVD88 NGVD29
#> Levels:  NAVD88 NGVD29
#>
#> $`country_id.levels(factor(y))`
#> [1] US
#> Levels: US
#>
#> $`state_id.levels(factor(y))`
#> [1] 12
#> Levels: 12
#>
#> $`county_id.levels(factor(y))`
#> [1] 57
#> Levels: 57
#>
#> $`aquifer.levels(factor(y))`
#> [1]                    Floridan aquifer system  Surficial aquifer system
#> Levels:  Floridan aquifer system Surficial aquifer system
#>
#> $`aquifer_formation.levels(factor(y))`
#> [1]                    Floridan Aquifer System  Nonartesian Sand Aquifer
#> Levels:  Floridan Aquifer System Nonartesian Sand Aquifer
#>
#> $`aquifer_type.levels(factor(y))`
#> [1]
#> [2] Confined single aquifer
#> [3] Mixed (confined and unconfined) multiple aquifers
#> [4] Unconfined single aquifer
#> 4 Levels:  ... Unconfined single aquifer
#>
#> $`construction_date.levels(factor(y))`
#>  [1] 197604   20011205 20011206 20011215 20020107 20020108 20020109 20020110
#>  [9] 20020111 20020112 20020119
#> 11 Levels: 197604 20011205 20011206 20011215 20020107 20020108 ... 20020119
#>
#> $`well_depth.levels(factor(y))`
#>  [1] 14.02 18.51 20.83 25.77 27.53 28.88 36.21 38    38.24 38.49 38.58 41.28
#> [13] 45    46.85 52    54    60    70    100   150   250
#> 21 Levels: 100 14.02 150 18.51 20.83 25.77 250 27.53 28.88 36.21 38 ... 70
#>
#> $`well_depth_unit.levels(factor(y))`
#> [1]    ft
#> Levels:  ft
#>
#> $`wellhole_depth.levels(factor(y))`
#>  [1] 19   23   28   33   38   39   43   45   51   54.7 55   55.5 61.7 70   100
#> [16] 150  250
#> Levels: 100 150 19 23 250 28 33 38 39 43 45 51 54.7 55 55.5 61.7 70
#>
#> $`wellhole_depth_unit.levels(factor(y))`
#> [1]    ft
#> Levels:  ft
#>
```

```
#> $`provider.levels(factor(y))`
#> [1] NWIS
#> Levels: NWIS

# Preview where locations of sampling sites are
# This requires the ggmap() package as well as a preconfigured Google Maps API key
# Please see: https://cran.r-project.org/web/packages/ggmap/readme/README.html for
# more information on the Google API key to properly generate the map
preview_locations(wqp)
#> Source : https://maps.googleapis.com/maps/api/staticmap?center=27.91,-82.35&zoom=10&size=640x640&sca
```

Sample Locations for WQP Data