

# ISCN3 visualization and report

*K Todd-Brown (ktoddbrown@gmail.com)*

*5/26/2017*

## Contents

<b>Observation time</b>	<b>2</b>
<b>Site locations</b>	<b>5</b>
Lat-lon map . . . . .	5
Sites with states but no lat-lon . . . . .	6
Sites with country but no lat-lon . . . . .	7
Site count by country map (include sites w/ lat-lon) . . . . .	9
<b>Summary of measurements</b>	<b>10</b>
Soil organic carbon stock . . . . .	11
Bulk density . . . . .	12
Carbon (mostly organic) . . . . .	14
Carbon (inorganic) . . . . .	15
Nitrogen . . . . .	15
pH . . . . .	16
Sand, clay, slit, and percent coarse fragment content . . . . .	17
Cation exchange capacity and base saturation . . . . .	18
Extractable bases . . . . .	20
Metal (Al, Fe, Mn) extracts . . . . .	21
Phosphorus . . . . .	23
Isotope . . . . .	24

The International Soil Carbon Network version 3 data base (ISCN3) is available here. Currently the Layers data is ingestion and summarized below (work is underway to also bring in profile level data). Much of the data is from NRCS (Sept 2014 version of the soil characterization data base) and methods can be referred to in SSL (Soil Survey Manual, R Burt) manual.

```
library(SoilDataR) #library(devtools); install_github("ktoddbrown/soilDataR")
library(ggplot2) #make pretty plots
library(plyr) #data management
library(dplyr)

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':
## 
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
```

```

library(knitr)
library(lubridate)

##
## Attaching package: 'lubridate'
## The following object is masked from 'package:plyr':
##   here
## The following object is masked from 'package:base':
##   date
library(tidyr)

#mapping librarys to help with global/regional plots
library(ggmap)
library(maps)

##
## Attaching package: 'maps'
## The following object is masked from 'package:plyr':
##   ozone
library(mapdata)
library(fiftystater)

ISCN3 <- SoilDataR::processData_ISCN3(dir='../../repoData/ISCN_3/')

## Warning: ISCN3 is a large data set and will take some time...
## Warning in processWorksheet_ISCN3(csvFile = sprintf("%s/Layers/
## ISCN_ALL_DATA_LAYER_C1_1-1.csv", : NAs introduced by coercion
## done!

```

## Observation time

```

location.df <- unique(ISCN3$field[, c('lat', 'long', 'observation_date',
                                         'state', 'country')])
location.df$observation_date <- as.character(location.df$observation_date)

location.df <- location.df %>%
  separate(observation_date, c('monthStr', 'dayStr', 'yearStr'),
            remove=FALSE, fill='left') %>%
  mutate(year=ifelse(as.numeric(yearStr) < 20, as.numeric(yearStr) + 2000,
                    ifelse(as.numeric(yearStr) < 100, as.numeric(yearStr) + 1900,
                           as.numeric(yearStr))),
         month=as.numeric(monthStr), day=as.numeric(dayStr)) %>%
  select(-contains('Str'))

location.df$obsDate <- ymd(paste(location.df$year, location.df$month, location.df$day, sep="-"))

```

```

## Warning: 3535 failed to parse.

location.df <- location.df[order(location.df$lat, location.df$long, location.df$obsDate),]
location.df$yrCut <- cut(location.df$year, c(1928, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2014), dig
#location.df$yrCut <- location.df$year
location.df$latCut <- cut(location.df$lat, seq(-90, 90, by=0.05))
location.df$longCut <- cut(location.df$long, seq(-180, 180, by=0.05))

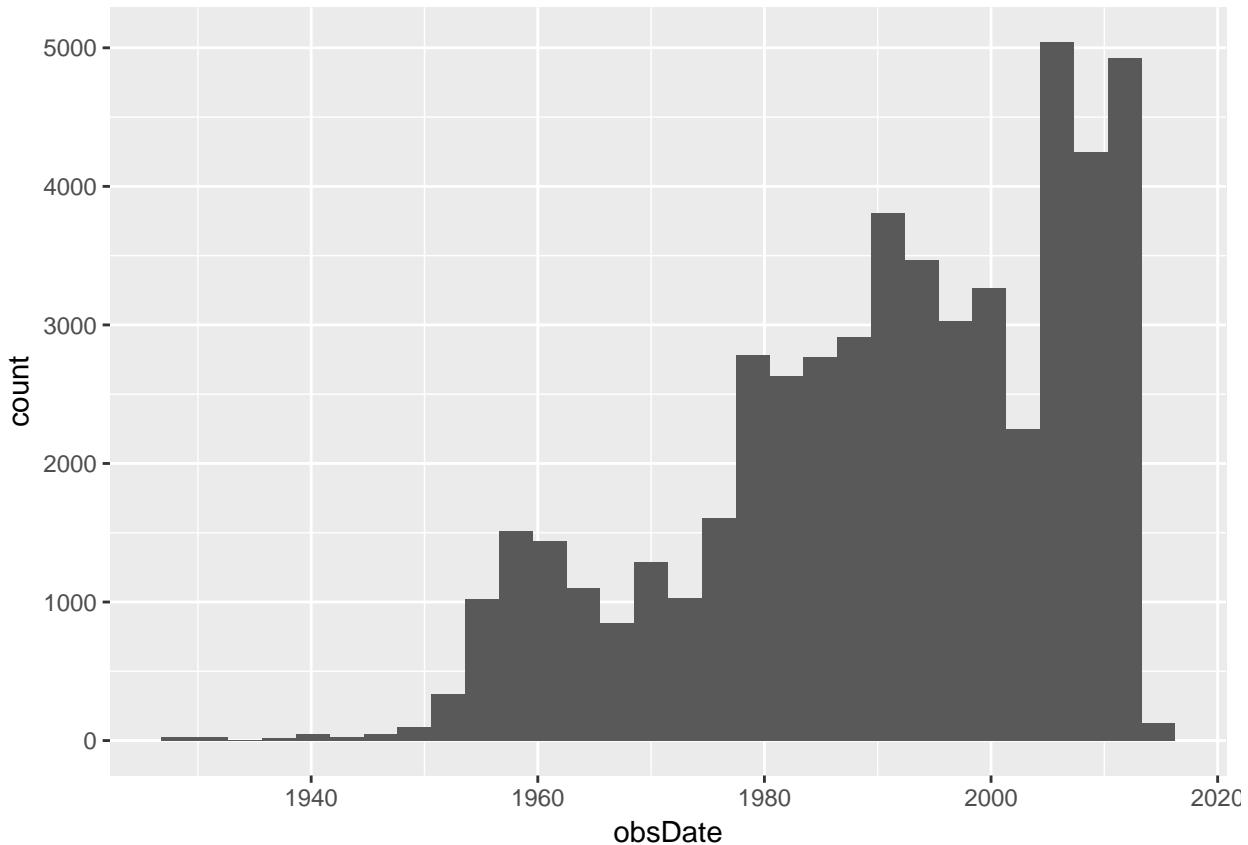
timeSpaceCounts <- location.df %>%
  mutate(inUSA = country %in% 'United States') %>%
  group_by(yrCut, latCut, longCut, inUSA) %>%
  tally

ggplot(location.df, aes(x=obsDate)) + geom_histogram()

```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 3535 rows containing non-finite values (stat\_bin).

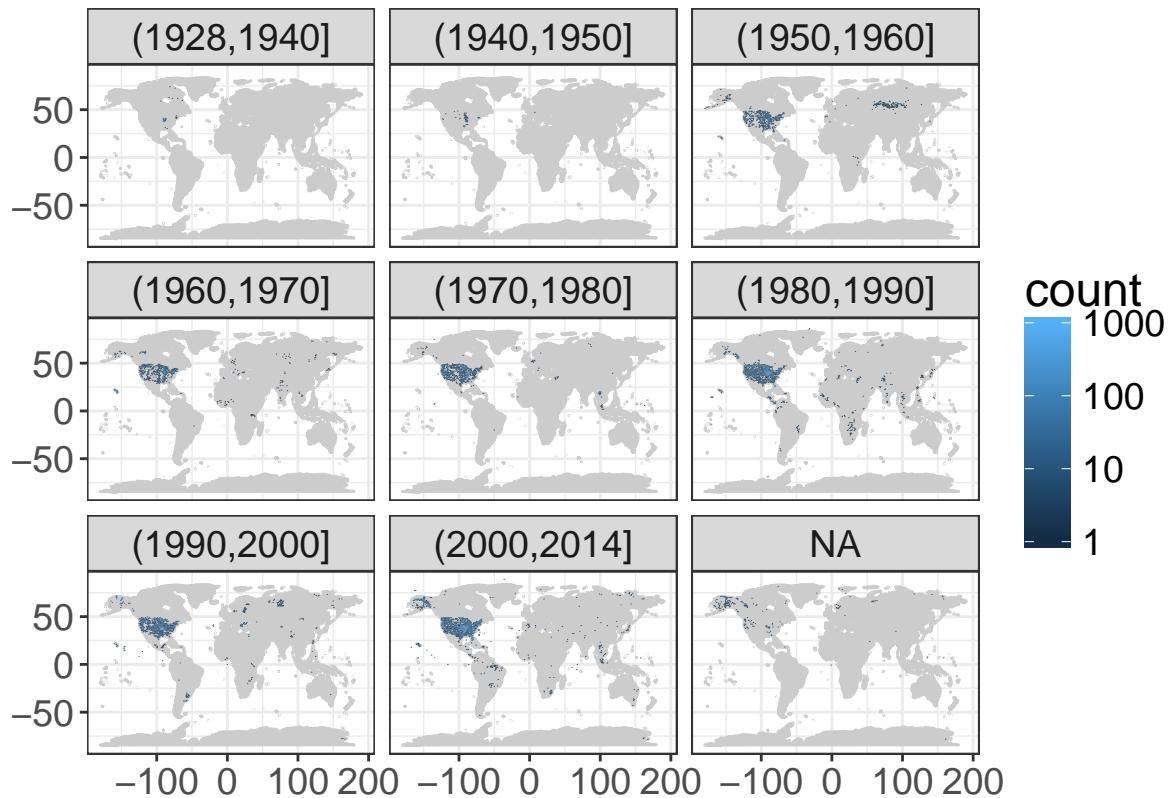


```

mapWorld <- borders("world", colour="gray80", fill="gray80") # create a layer of borders
#ggplot() + mapWorld
ggplot(location.df) +
  mapWorld +
  geom_hex(aes(x=long, y=lat), bins=200) +
  scale_fill_gradient(trans='log10') +
  theme_bw() +
  theme(text=element_text(size=18)) +
  labs(x='', y='') +
  facet_wrap(~yrCut)

```

```
## Warning: Removed 3822 rows containing non-finite values (stat_binhx).
```

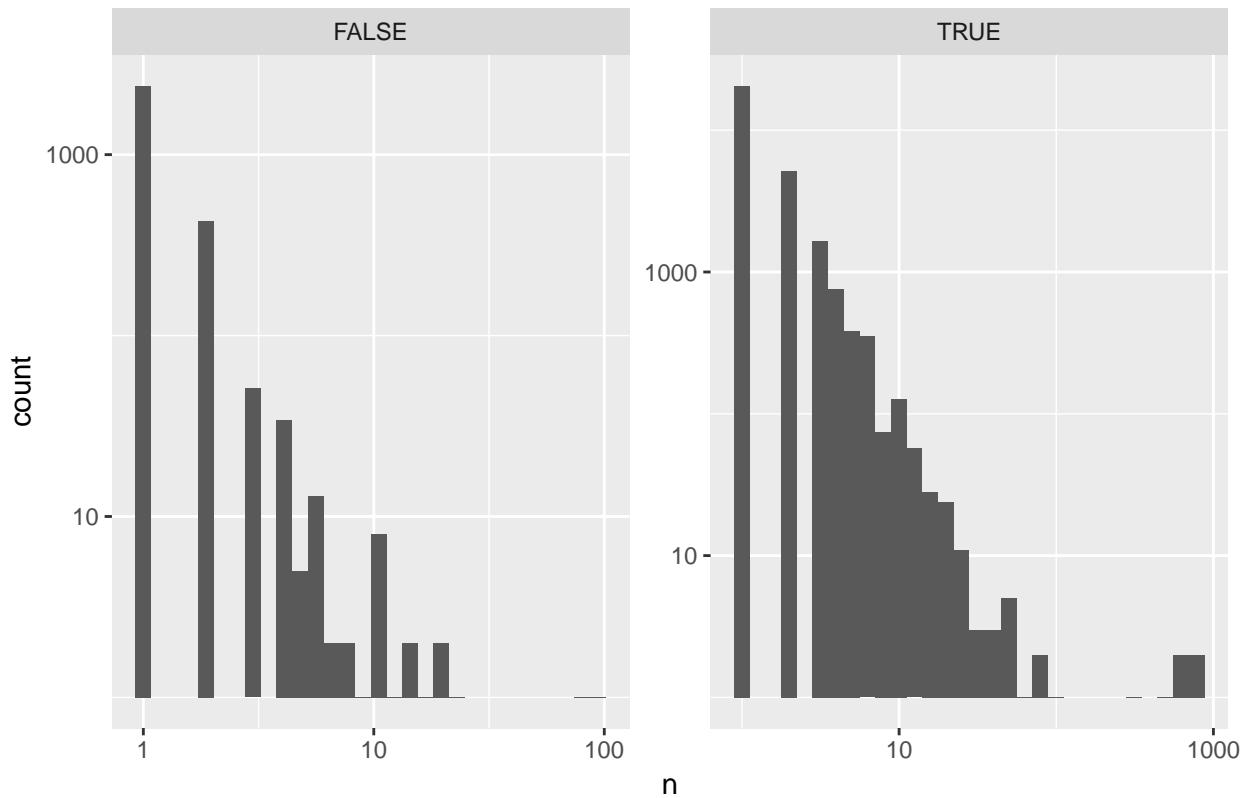


```
#ggplot(location.df, aes(x=long, y=lat)) + geom_point(alpha=0.1)

ggplot(timeSpaceCounts, aes(x=n)) + geom_histogram() +
  scale_x_log10() + scale_y_log10() + facet_wrap(~inUSA, scales='free') +
  labs(title='Close space-time sites by USA status')

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Transformation introduced infinite values in continuous y-axis
## Warning: Removed 21 rows containing missing values (geom_bar).
```

## Close space–time sites by USA status

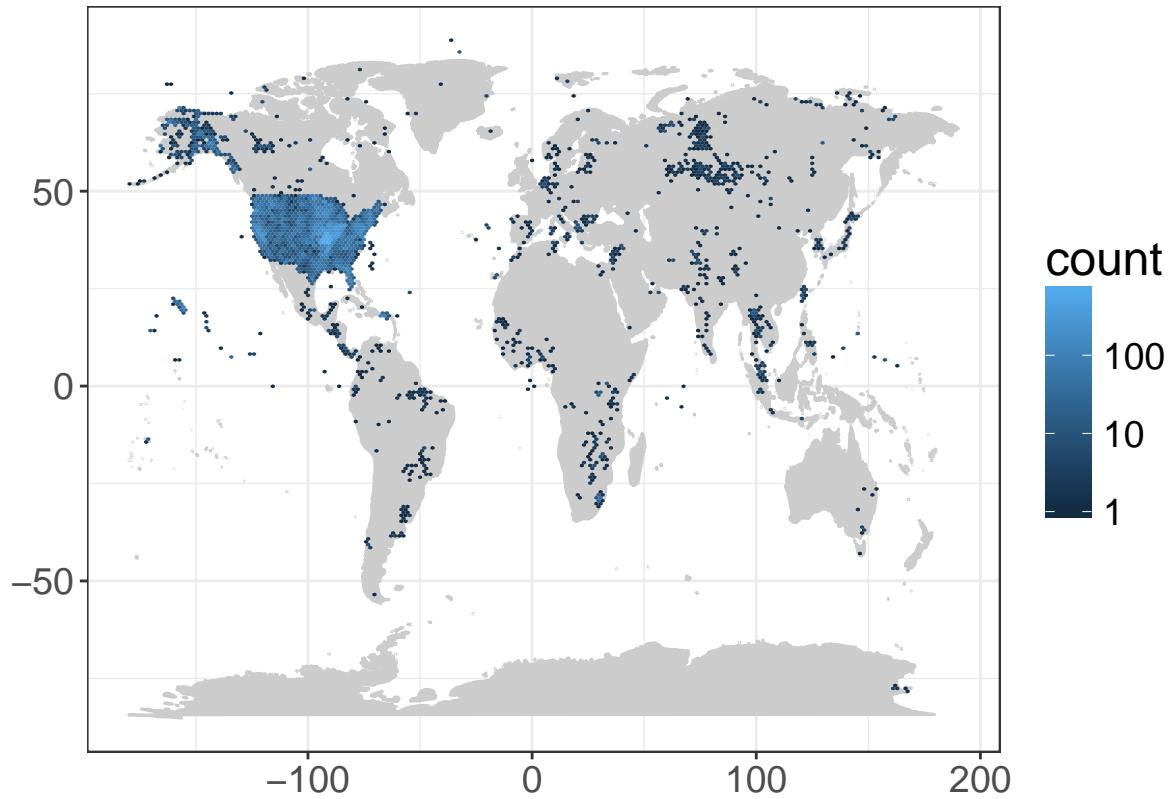


## Site locations

### Lat-lon map

```
mapWorld <- borders("world", colour="gray80", fill="gray80") # create a layer of borders
#ggplot() + mapWorld
ggplot(unique(ISCN3$field[, c('lat', 'long')])) +
  mapWorld +
  geom_hex(aes(x=long, y=lat), bins=200) +
  scale_fill_gradient(trans='log10') +
  theme_bw() +
  theme(text=element_text(size=18)) +
  labs(x='', y='')

## Warning: Removed 1 rows containing non-finite values (stat_binhex).
```



### Sites with states but no lat-lon

```

noLatLon <- unique(subset(ISCN3$field, is.na(lat+long))$fieldID)

stateCounts <- ddply(subset(ISCN3$field, is.na(lat+long)), c('state'),
                      summarize, count=length(unique(fieldID)))
stateCounts$state <- tolower(stateCounts$state)

cat('Following are not considered states:\n')

## Following are not considered states:
kable(subset(stateCounts, ! state %in% fifty_states$id))

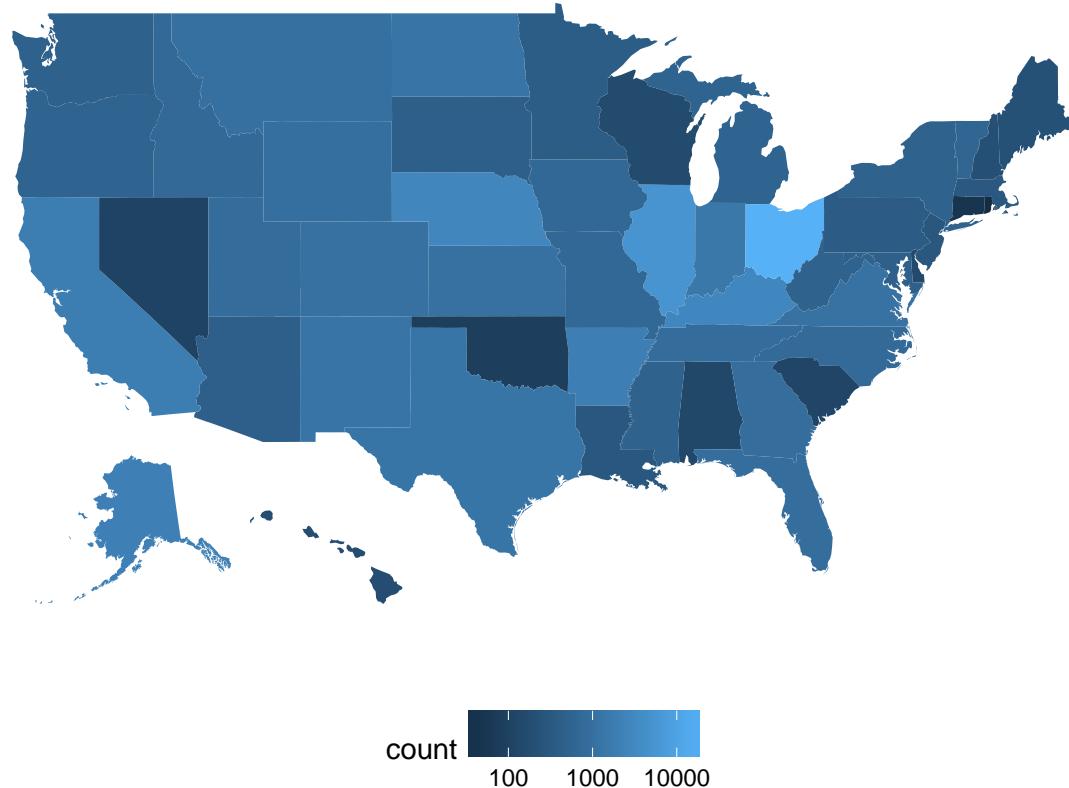
```

	state	count
1		4720
10	guam	6
15	kosrae	3
17	manicaland	17
18	midlands	6
24	northern mariana islands	14
26	pohnpei	2
27	puerto rico	404
31	unknown	3745
33	virgin islands	25
46	cataluna	2
62	'ajman	13

state	count
63 chiapas	6
64 national lab	748
65 tsumeb	29

```
stateCounts <- subset(stateCounts, state %in% fifty_states$id)

ggplot(stateCounts, aes(map_id = state)) +
  geom_map(aes(fill=count), map=fifty_states) +
  expand_limits(x = fifty_states$long, y = fifty_states$lat) +
  coord_map() +
  scale_x_continuous(breaks = NULL) +
  scale_y_continuous(breaks = NULL) +
  scale_fill_gradient(trans='log10') +
  labs(x = "", y = "") +
  theme(legend.position = "bottom",
        panel.background = element_blank())
```



## Sites with country but no lat-lon

```
countryCounts <- ddply(subset(ISCN3$field, is.na(lat+long)), 'country', summarize, count=length(unique
names(countryCounts)[1] <- 'region'
countryCounts$region <- as.character(countryCounts$region)
countryCounts$region[grep('United States', countryCounts$region)] <- 'USA'

map.world <- map_data(map="world")
```

```

cat('These countries are unknown to the map\n')

## These countries are unknown to the map
kable(subset(countryCounts, ! region %in% map.world$region))

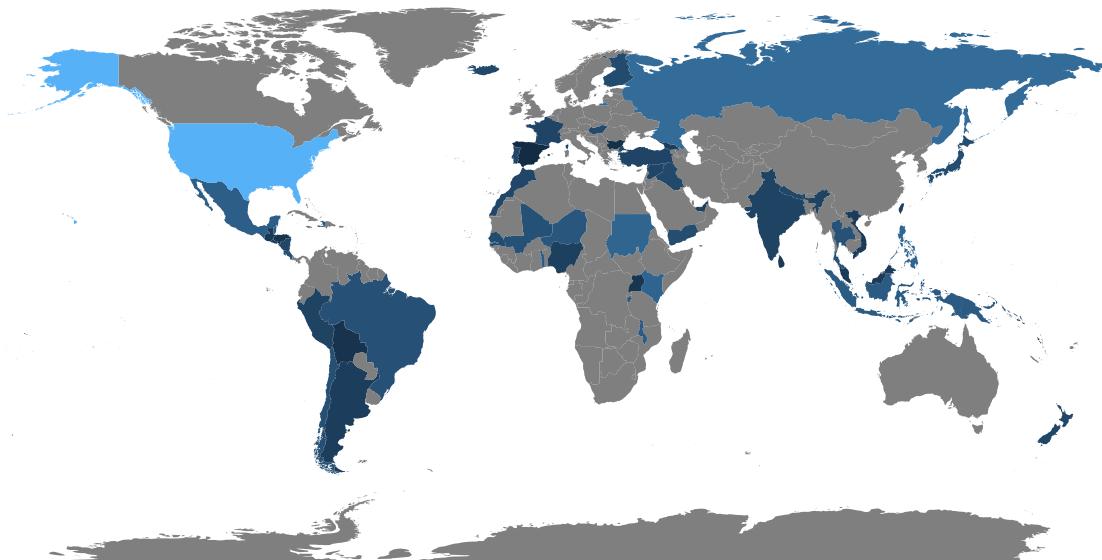
```

	region	count
1		4720
6	Congo (Democratic Republic of the)	11
9	Federated States of Micronesia	11
17	Korea, Republic of	4
37	Unknown	2596
50	Korea, Democratic People's Republic of	67

```

missingCountries <- data.frame(region=setdiff(map.world$region, countryCounts$value), count=NA)
countryCounts <- rbind(countryCounts, missingCountries)
#map.world <- merge(map.world, countryCounts, all=TRUE)
ggplot(countryCounts, aes(map_id = region)) +
  geom_map(aes(fill=count), map=map.world) +
  expand_limits(x = map.world$long, y = map.world$lat) +
  coord_cartesian(xlim=c(-180,180), ylim=c(-90,90)) +
  scale_x_continuous(breaks = NULL) +
  scale_y_continuous(breaks = NULL) +
  scale_fill_gradient(trans='log10') +
  labs(x = "", y = "") +
  theme(legend.position = "bottom",
        panel.background = element_blank())

```



7316 sites with no known country nor lat-lon.

### Site count by country map (include sites w/ lat-lon)

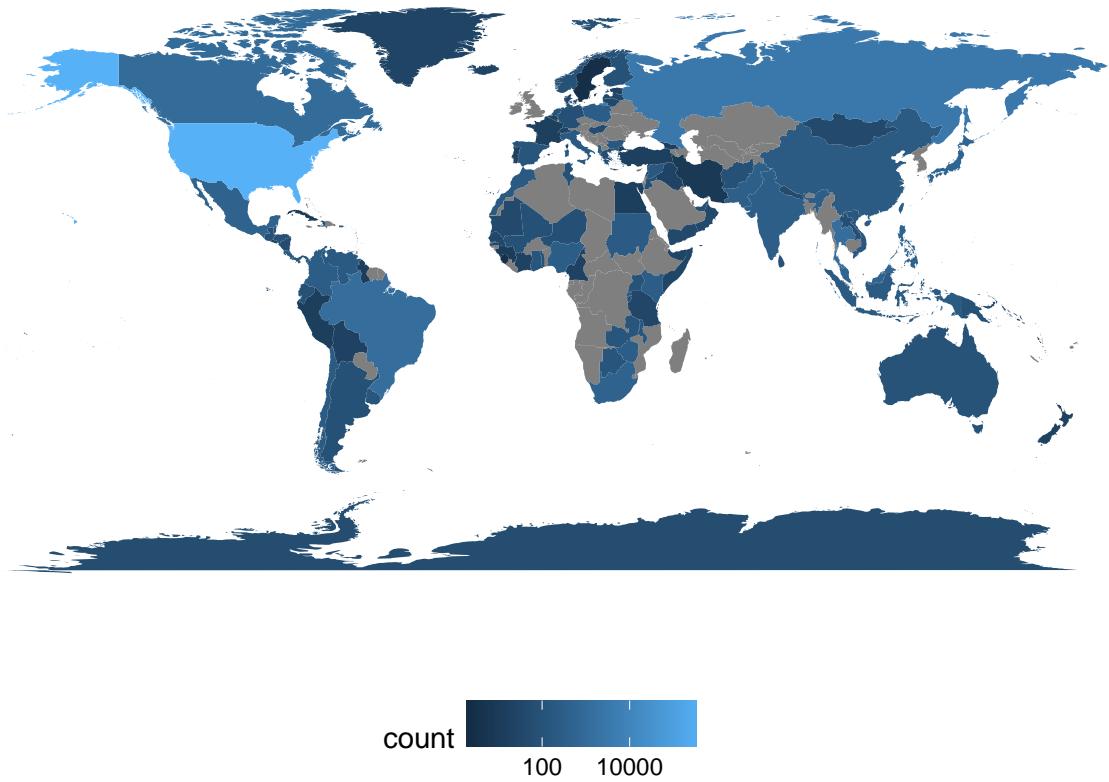
```
countryCounts <- ddply(ISCN3$field, 'country', summarize, count=length(unique(fieldID)))
names(countryCounts)[1] <- 'region'
countryCounts$region <- as.character(countryCounts$region)
countryCounts$region[grep('United States', countryCounts$region)] <- 'USA'

map.world <- map_data(map="world")

missingCountries <- data.frame(region=setdiff(map.world$region, countryCounts$region), count=NA)
setdiff(countryCounts$region, map.world$region)

## [1] ""
## [2] "Congo (Democratic Republic of the)"
## [3] "Dahomey"
## [4] "Federated States of Micronesia"
## [5] "Korea, Republic of"
## [6] "Portuguese Guinea"
## [7] "Tasmania"
## [8] "Unknown"
## [9] "Yugoslavia"
## [10] "Korea, Democratic People's Republic of"

countryCounts <- rbind(countryCounts, missingCountries)
#map.world <- merge(map.world, countryCounts, all=TRUE)
ggplot(countryCounts, aes(map_id = region)) +
  geom_map(aes(fill=count), map=map.world) +
  expand_limits(x = map.world$long, y = map.world$lat) +
  coord_cartesian(xlim=c(-180,180), ylim=c(-90,90)) +
  scale_x_continuous(breaks = NULL) +
  scale_y_continuous(breaks = NULL) +
  scale_fill_gradient(trans='log10') +
  labs(x = "", y = "") +
  theme(legend.position = "bottom",
        panel.background = element_blank())
```



## Summary of measurements

The following variables are listed in

```
table(ISCN3$measurement$type)
```

	bd_samp	bd_tot	bd_whole
##	6	5	2
##	bd_other	c_tot	oc
##	14	59	20
##	loi	n_tot	c_to_n
##	9	1	0
##	soc	ph_cacl	ph_h2o
##	130	2	4
##	ph_other	caco3	sand_tot_psa
##	2	1	1
##	silt_tot_psa	clay_tot_psa	cat_exch
##	1	1	1
##	wpg2	al_dith	al_ox
##	6	1	1
##	al_other	fe_dith	fe_ox
##	1	1	1
##	fe_other	mn_dith	mn_ox
##	1	1	1
##	mn_other	ca_al	ca_ext
##	1	0	1
##	k_ext	mg_ext	na_ext

```

##          1          1          1
##    base_sum      cec_sum     ecec
##          1          1          1
##      bs      bs_sum     h_ext
##          1          1          0
##    zn_ext      p_bray     p_ox
##          0          1          1
##    p_meh      p_other   root_weight
##          0          1          0
##      15n      13c      14c
##          1          1          1
##    14c_sigma    14c_age  14c_age_sigma
##          0          0          0
## fraction_modern fraction_modern_sigma
##          0          0

```

All measurements are on oven dry soils and in mass percent where percent noted.

## Soil organic carbon stock

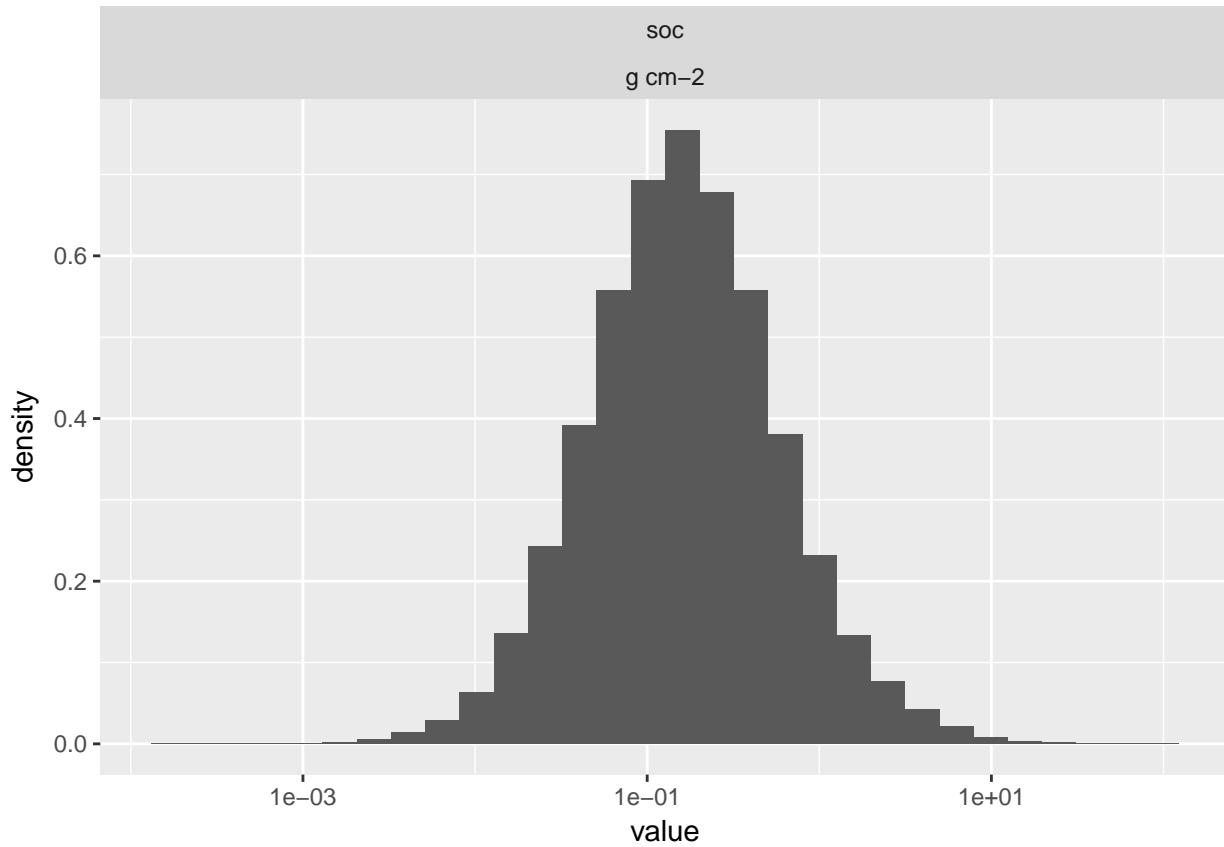
- *soc* calculated soil carbon stock of layer/profile

```

ggplot(merge(subset(ISCN3$sample, grep('soc', measurementID) & value > 0),
            ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density..)) +
  scale_x_log10() +
  facet_wrap(type~unit)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```



```
summary(subset(ISCN3$sample, grep('soc', measurementID) & (value < 0)))
```

##	fieldID	measurementID	value	unit
##	GCPUs* 5.0rge,abm :1	soc_01 :2	Min. :-998	g cm-2 :2
##	GCPUs* 5.0rgi,abm :1	13c_01 :0	1st Qu.:-998	% :0
##	00015-1-1 :0	14c_01 :0	Median :-998	cmol H+ kg-1:0
##	00015-1-2 :0	15n_01 :0	Mean :-998	g cm-3 :0
##	00015-1-3 :0	al_dith_01 :0	3rd Qu.:-998	meq 100g-1 :0
##	00015-1-4 :0	al_other_01:0	Max. :-998	percent :0
##	(Other) :0	(Other) :0		(Other) :0

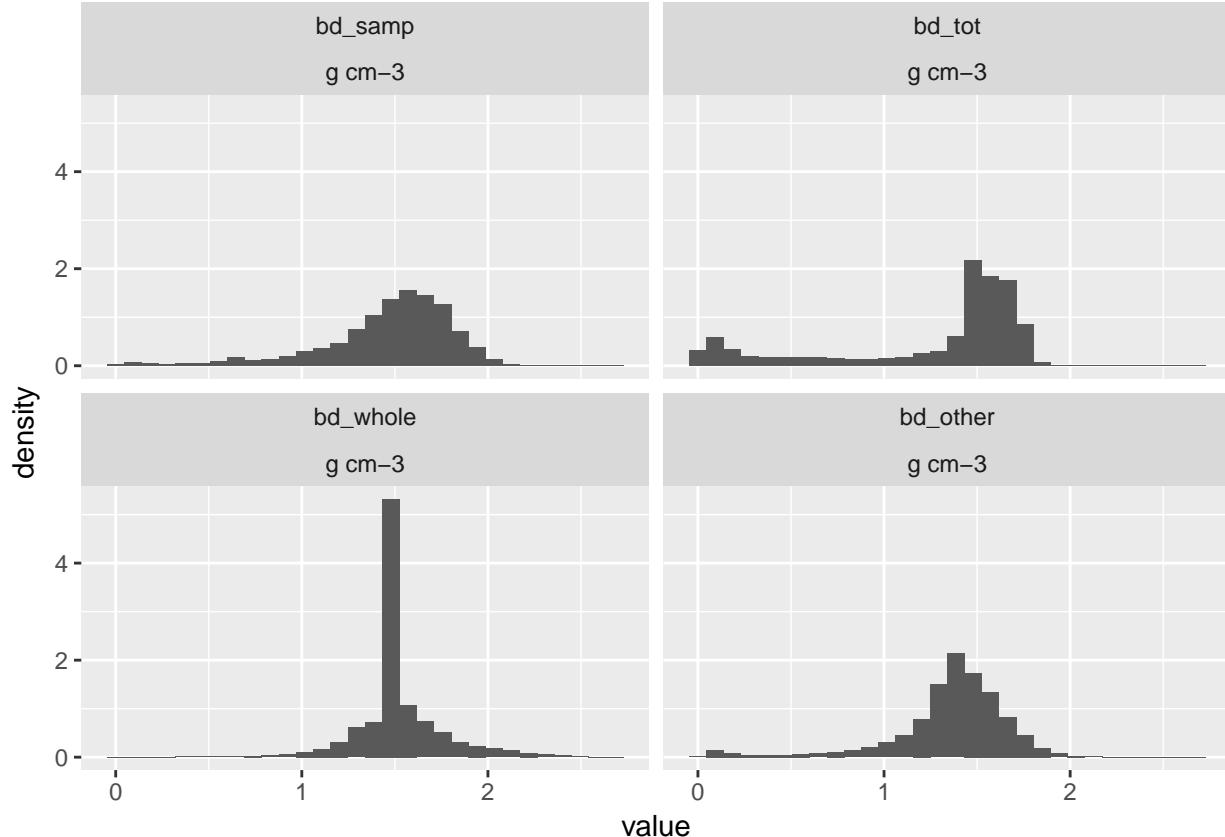
## Bulk density

- *bd\_samp* is the bulk density of the <2mm (fine earth) fraction, in which the mass is expressed on an oven-dry (105 deg) basis. This is the best form of Db for the most uses.
- *bd\_tot* is the whole soil bulk density, includes fine earth fraction and rocks.
- *bd\_whole* is the bulk density of the whole soil (coarse frags, if present, included), expressed on a 1/3 bar moisture content (a little drier than typical field moist, but not oven-dry). It is collected by the clod method. For a variety of reasons, including the presence of water, possible coarse frags, and the tendency of clods to cleave along pore spaces, this will overestimate Db relative to *bd\_samp*. The SSL standard method here is 4A1d if you want to look it up.
- *bd\_other* is, for data contributed by NRCS SSL, the bulk density of the fine earth fraction, but expressed on a field moist condition. It is collected by a volumetric core, but given the presence of some water, will probably usually be an overestimate relative to the *bd\_samp*. This corresponds to the SSL standard method 4A3a. For datasets contributed by sources other than NRCS, *bd\_other* is anyone's guess. Hopefully specified in the associated *bd\_method* variable, or the metadata sheet for that dataset.

Reasonable BD minimum 0.03 or lower and maximum 2.7 g cm<sup>-3</sup>.

```
plot.df <- merge(subset(ISCN3$sample, grep('^bd_', measurementID) &
                           value > 0 & value < 2.7), ISCN3$measurement)
ggplot(plot.df) +
  geom_histogram(aes(x=value, y=..density..)) +
  #scale_x_log10() +
  facet_wrap(type~unit)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
summary(subset(ISCN3$sample, grep('^bd_', measurementID) &
                           (value <= 0 | value > 2.7)))
```

##	fieldID	measurementID	value	unit
##	00P02751 : 2	bd_samp_06 : 17	Min. : -998.00	g cm <sup>-3</sup> : 48
##	BF84-1-F : 2	bd_tot_01 : 10	1st Qu.: 0.00	% : 0
##	FP3C_4_17: 2	bd_whole_02: 8	Median : 0.00	cmol H <sup>+</sup> kg <sup>-1</sup> : 0
##	HCCN 5.47: 2	bd_samp_01 : 5	Mean : -12.85	g cm <sup>-2</sup> : 0
##	TFBC 6.9 : 2	bd_other_07: 2	3rd Qu.: 3.62	meq 100g <sup>-1</sup> : 0
##	TFBC 7.7 : 2	bd_tot_05 : 2	Max. : 1230.00	percent : 0
##	(Other) : 36	(Other) : 4		(Other) : 0

```
nrow(plot.df)
```

```
## [1] 413921
```

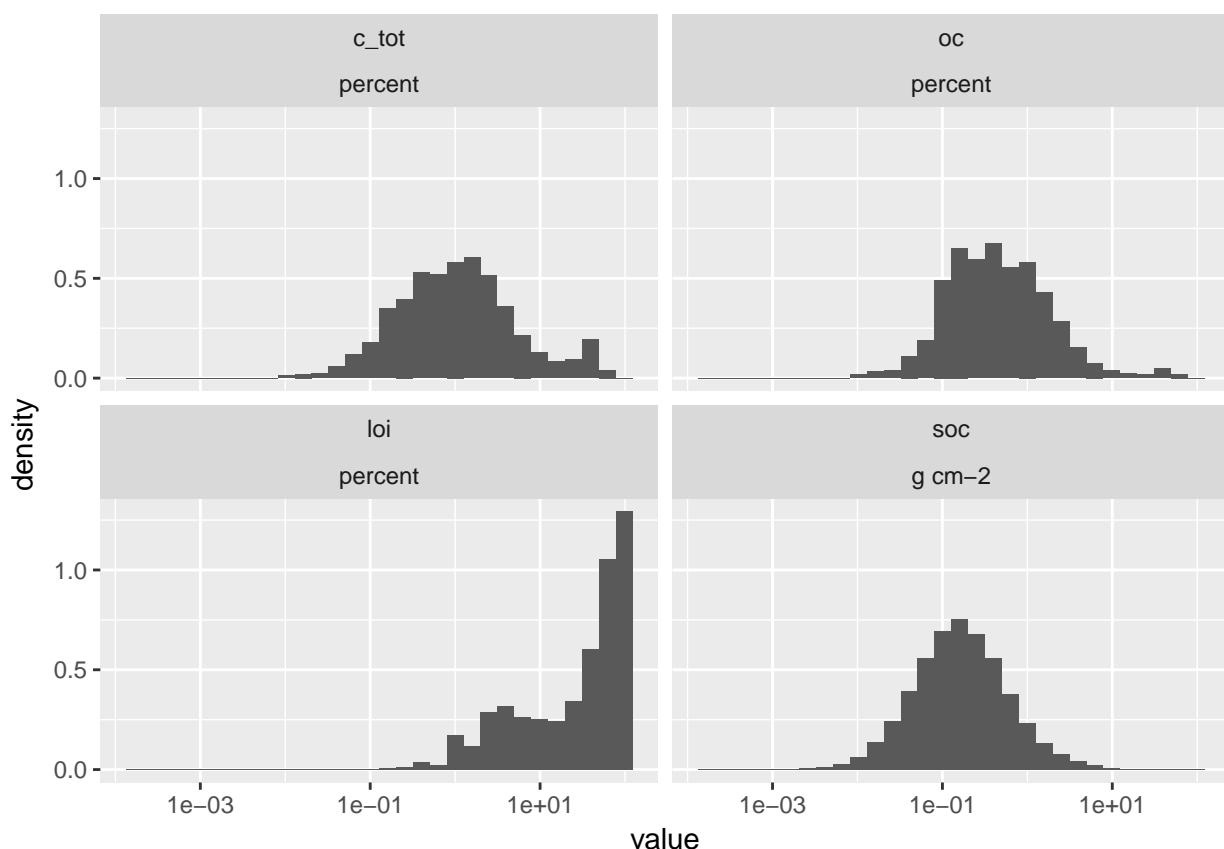
Number of samples for BD 413921.

## Carbon (mostly organic)

- *c\_tot* carbon concentration in a dry combustion analysis, sometimes include inorganic carbon
- *oc* organic carbon concentration from either acidified (to remove carbonates) and then ran on dry combustion analyzer, or (more common) wet oxidation ‘Walkly-Black’ dicromate oxidation, chemical digest for organic carbon; old methods 1930-1990s tends to over estimate it in certain soils. pH below 7 *c\_tot* generally == *oc* otherwise you might have carbonates
- *loi* loss on ignition uncorrected for C:organics, generally close to half for O-horizon, C about a third of the LOI for lower horizons.

```
ggplot(merge(subset(ISCN3$sample, grep('^(c_|oc|loi|soc)', measurementID) & value > 0),
            ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density..)) +
  scale_x_log10() +
  facet_wrap(type~unit)
```

## `stat\_bin()` using `bins = 30` . Pick better value with `binwidth` .



```
summary(subset(ISCN3$sample, grep('^(c_|oc|loi)', measurementID) & (value < 0)))
```

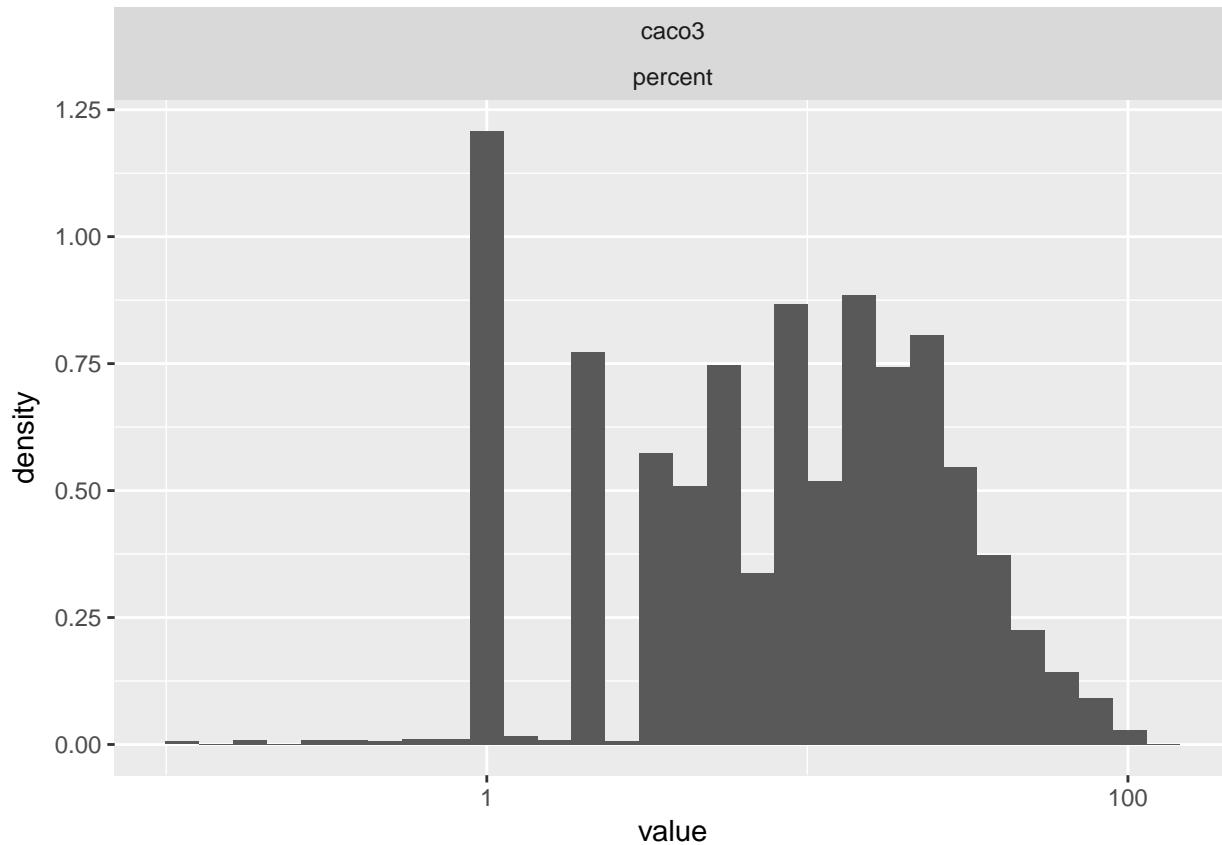
```
##      fieldID      measurementID      value          unit
## 02N05882 :1    loi_08      :1   Min.   :-102.00  percent     :2
## 72C00098 :1    oc_19      :1  1st Qu.: -76.53   %        :0
## 00015-1-1:0  13c_01      :0 Median  : -51.07 cmol H+ kg-1:0
## 00015-1-2:0  14c_01      :0 Mean    : -51.07 g cm-2     :0
## 00015-1-3:0  15n_01      :0 3rd Qu.: -25.61 g cm-3     :0
## 00015-1-4:0 al_dith_01:0 Max.    :  -0.14 meq 100g-1 :0
## (Other)    :0 (Other)    :0
```

## Carbon (inorganic)

Percent inorganic carbon in a sample.

```
ggplot(merge(subset(ISCN3$sample, grepl('caco3', measurementID) & value > 0),
             ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density..)) +
  scale_x_log10() +
  facet_wrap(type~unit)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
summary(subset(ISCN3$sample, grepl('caco3', measurementID) & (value <= 0)))
```

```
##      fieldID      measurementID      value      unit
## 00P00005: 1    caco3_01   :25036  Min.   :0  percent   :25036
## 00P00006: 1    13c_01     : 0    1st Qu.:0  %        : 0
## 00P00007: 1    14c_01     : 0    Median :0  cmol H+ kg-1: 0
## 00P00014: 1    15n_01     : 0    Mean    :0  g cm-2   : 0
## 00P00015: 1    al_dith_01 : 0    3rd Qu.:0  g cm-3   : 0
## 00P00016: 1    al_other_01: 0    Max.   :0  meq 100g-1 : 0
## (Other) :25030 (Other)     : 0
```

## Nitrogen

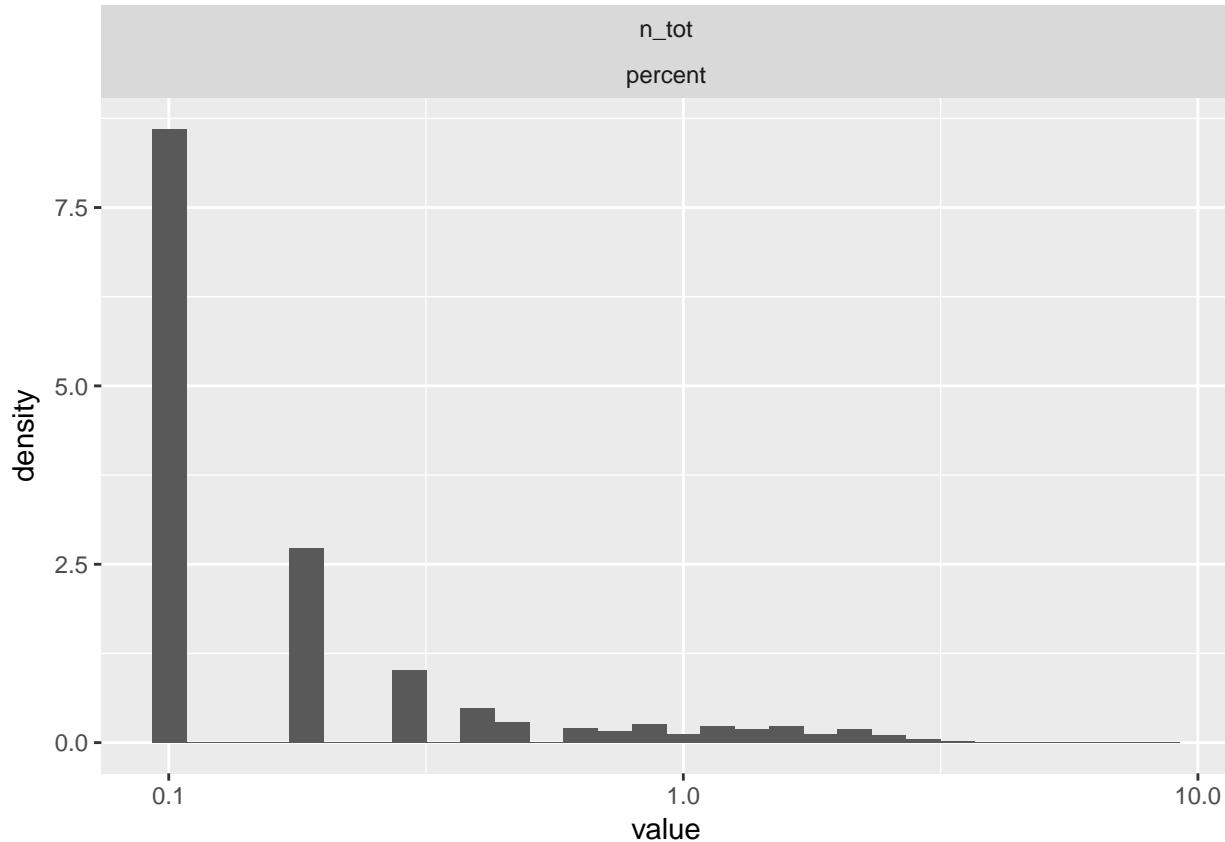
- $n_{tot}$  most N is organic mass percent dry weight, probably estimated by dry combustion (Dumas methods), sometimes also Kjeldahl digestion converts N to nitrate and then measures nitrate.

```

ggplot(merge(subset(ISCN3$sample, grep('n_', measurementID) & value > 0),
            ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density...)) +
  scale_x_log10() +
  facet_wrap(type~unit)

```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## pH

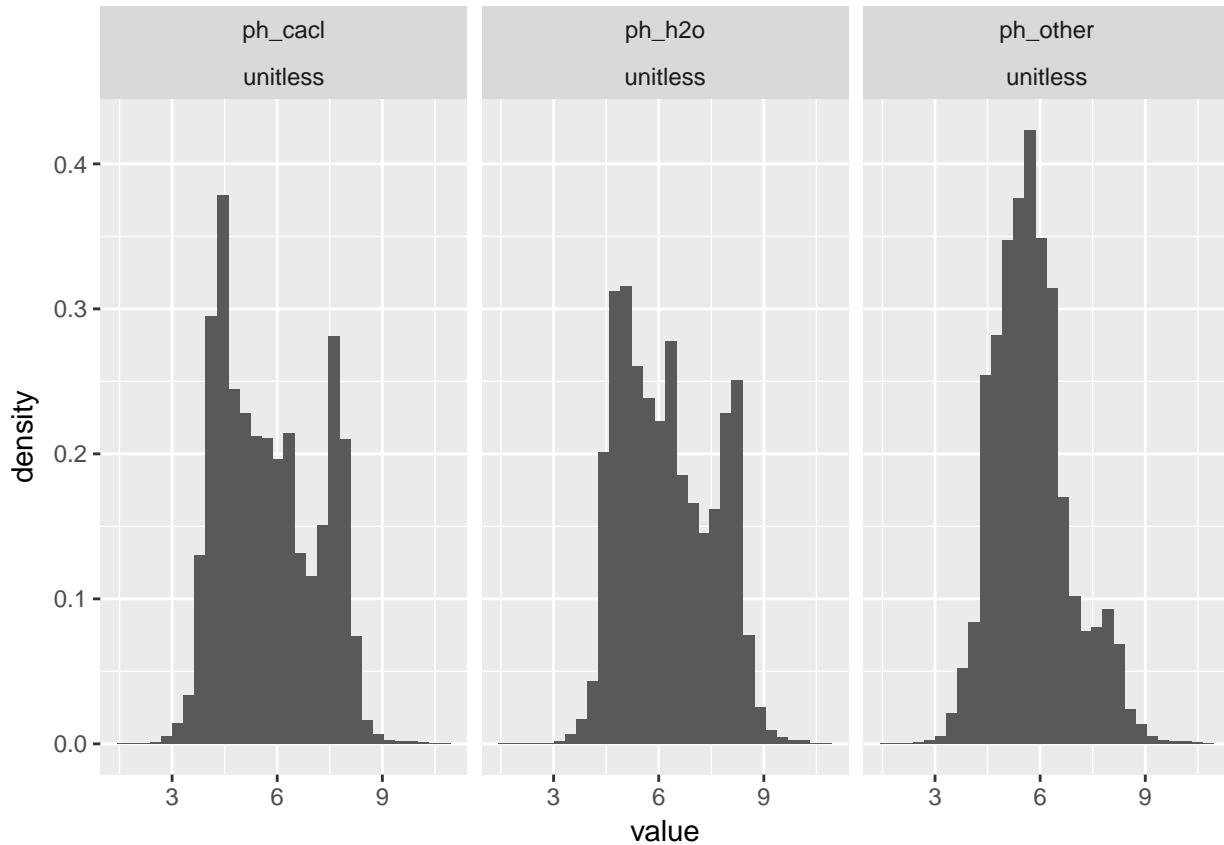
- *ph\_cacl* 1:2 Soil-CaCl<sub>2</sub> suspension
- *ph\_h2o* 1:1 Soil-water suspension
- *ph\_other* non-standard catch all

```

ggplot(merge(subset(ISCN3$sample, grep('ph_', measurementID) &
                           value > 1.5 & value < 11),
            ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density...)) +
  #scale_x_log10() +
  facet_wrap(type~unit)

```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



```
summary(subset(ISCN3$sample, grepl('^ph_', measurementID) & (value <= 1.5 | value > 11)))

##      fieldID      measurementID      value          unit
## MU002991 : 2 ph_h2o_01 :23   Min. :-7.700  unitless    :40
## MU008274 : 2 ph_other_01: 7  1st Qu.: 0.000   %           : 0
## 4010002.1: 1 ph_h2o_02 : 6 Median : 0.100 cmol H+ kg-1: 0
## 4010002.2: 1 ph_cacl_01 : 4 Mean   : 5.303 g cm-2     : 0
## 4010002.3: 1 13c_01     : 0 3rd Qu.: 0.700 g cm-3     : 0
## 4010002.4: 1 14c_01     : 0 Max.   :87.100 meq 100g-1 : 0
## (Other)   :32 (Other)    : 0                               (Other)    : 0

#d_ply(subset(ISCN3$sample, grepl('^ph_', measurementID) & (value <= 1.5 | value > 11)),
#       'measurementID', function(xx){print(summary(xx))})
```

## Sand, clay, slit, and percent coarse fragment content

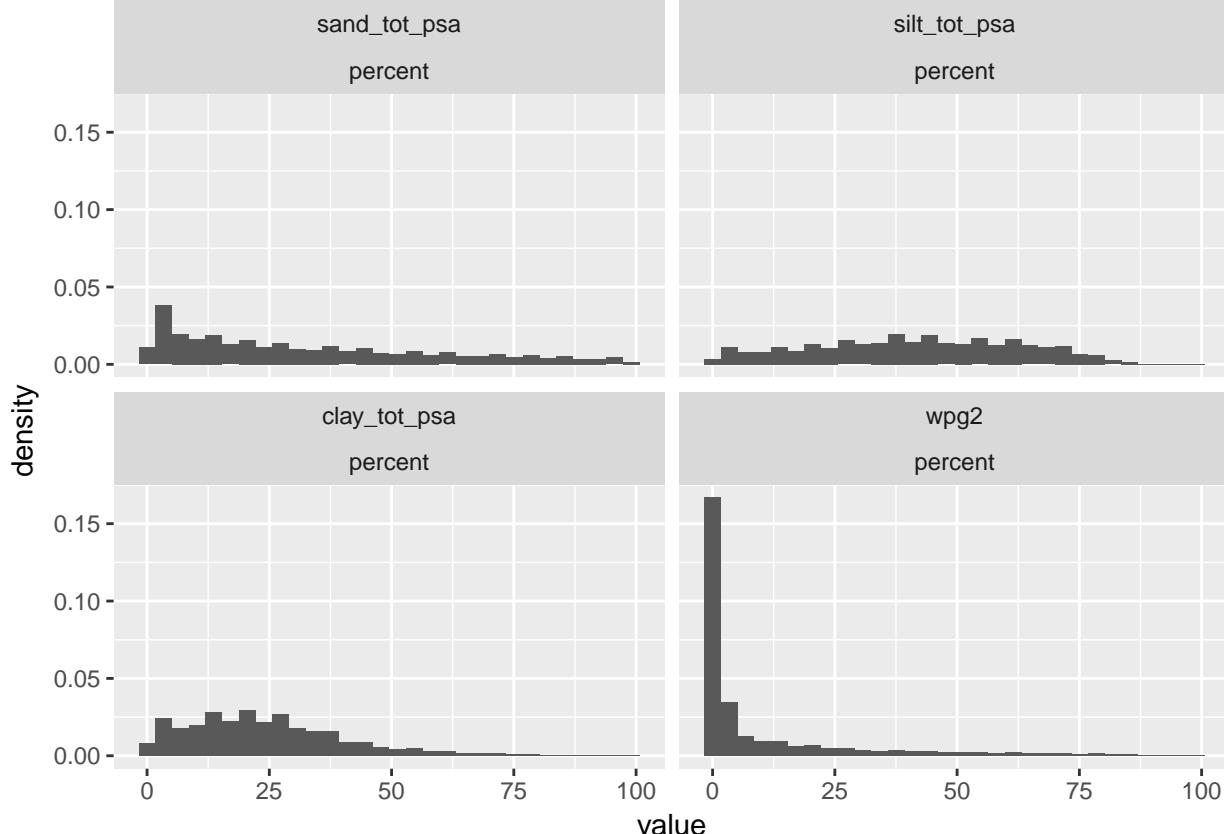
Partical size analysis. Methods might not be completely annotated.

- *wpg2* coarse fragment (>2mm) content by weight
- *sand\_tot\_psa* percentage of sand
- *clay\_tot\_psa* percentage of clay
- *silt\_tot\_psa* percentage of silt

```
ggplot(merge(subset(ISCN3$sample, grepl('(wpg|tot_psa)', measurementID) &
                           value < 100 & value >= 0 ),
            ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density..)) +
```

```
#scale_x_log10() +
facet_wrap(type~unit)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
summary(subset(ISCN3$sample, grep('wpg|tot_psa', measurementID) &
               (value < 0 | value > 100)))
```

##	fieldID	measurementID	value	unit
##	40A40506: 2	wpg2_01 :21	Min. : -62.0	percent :42
##	04N00472: 1	silt_tot_psa_01:11	1st Qu.: -2.0	% : 0
##	04N00476: 1	clay_tot_psa_01: 9	Median : 101.0	cmol H+ kg-1: 0
##	06N00201: 1	sand_tot_psa_01: 1	Mean : 318.0	g cm-2 : 0
##	06N04190: 1	13c_01 : 0	3rd Qu.: 121.2	g cm-3 : 0
##	06N04193: 1	14c_01 : 0	Max. : 4229.0	meq 100g-1 : 0
##	(Other) :35	(Other) : 0		(Other) : 0

## Cation exchange capacity and base saturation

- *cat\_exch* cation exchange capacity [[TODO: Ask Luke N about this variable]]
- *cec\_sum* cation exchange capacity (sum of cations)
- *ecec* effective cation exchange capacity
- *bs* base saturation (NH<sub>4</sub>OAc at pH 7.0)
- *bs\_sum* base saturation (sum of cations at pH 8.2)

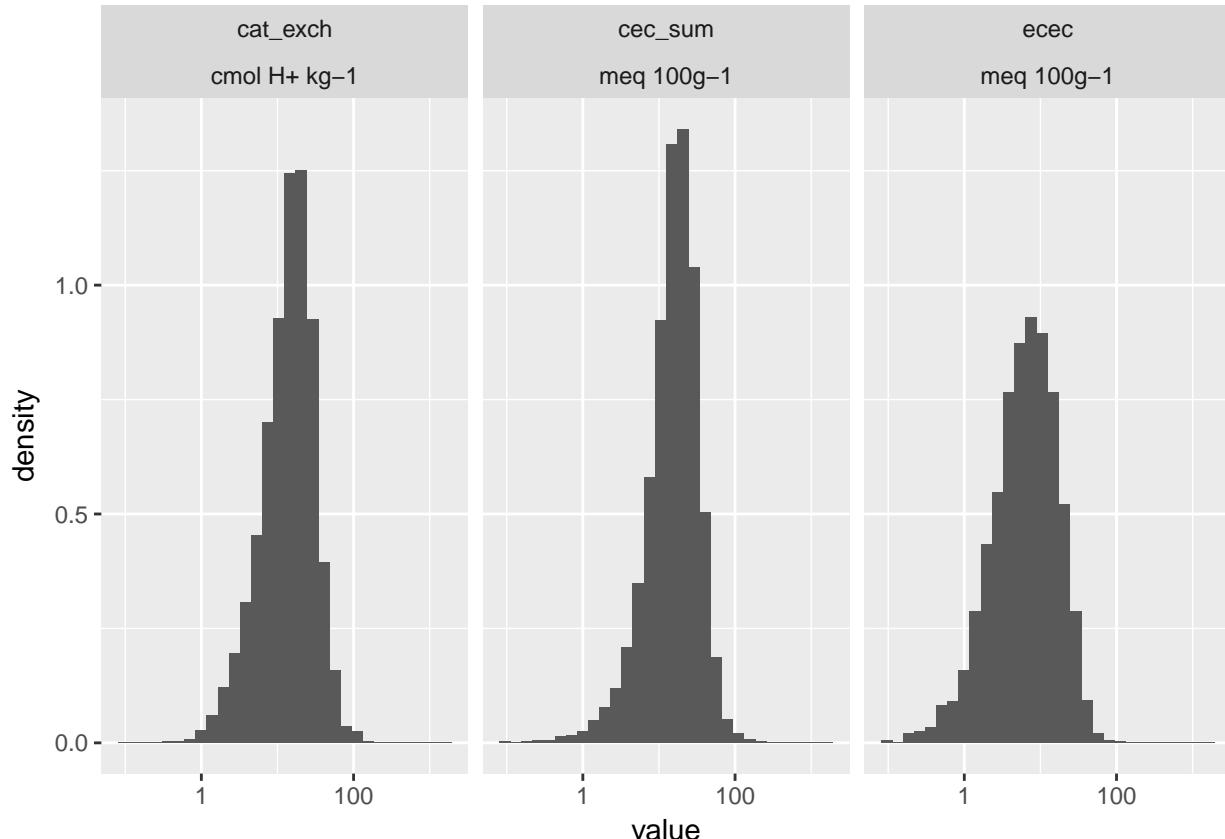
```
ggplot(merge(subset(ISCN3$sample, grep('cat_exch|cec_sum|ecec', measurementID)& (value > 0)),
            ISCN3$measurement)) +
```

```

geom_histogram(aes(x=value, y=..density..)) +
scale_x_log10() +
facet_wrap(type~unit)

```

## `stat\_bin()` using `bins = 30` . Pick better value with `binwidth` .



```
summary(subset(ISCN3$sample, grep('cat_exch|cec_sum|ecec', measurementID) & (value <= 0)))
```

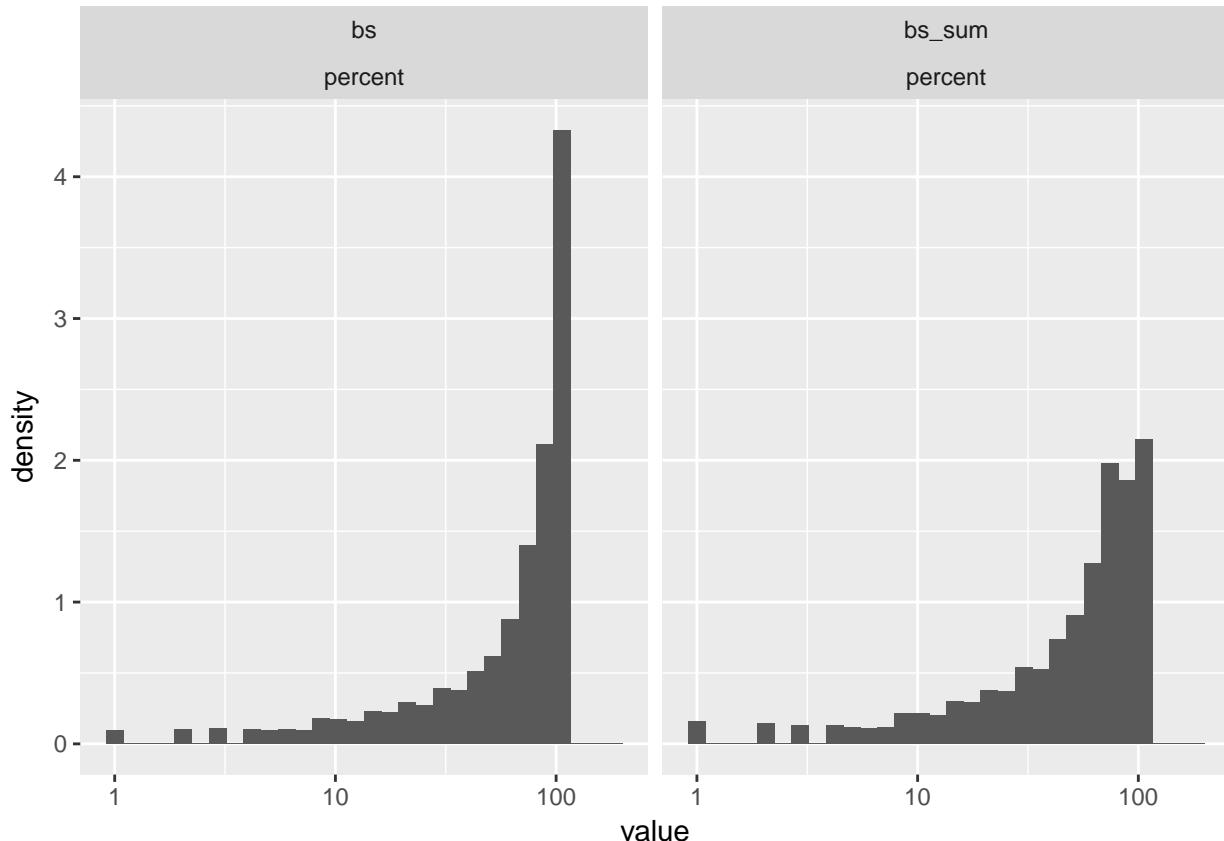
	fieldID	measurementID	value	unit
##	MU000161:	2 ecec_01	:382 Min.	:0 meq 100g-1 :649
##	MU000162:	2 cec_sum_01	:267 1st Qu.:	:0 cmol H+ kg-1: 1
##	MU000163:	2 cat_exch_01:	1 Median :	:0 % : 0
##	MU000502:	2 13c_01	: 0 Mean :	:0 g cm-2 : 0
##	MU000599:	2 14c_01	: 0 3rd Qu.:	:0 g cm-3 : 0
##	MU001641:	2 15n_01	: 0 Max. :	:0 percent : 0
##	(Other) :638	(Other)	: 0	(Other) : 0

```

ggplot(merge(subset(ISCN3$sample, grep('bs|bs_sum', measurementID)& (value > 0)),
ISCN3$measurement)) +
geom_histogram(aes(x=value, y=..density..)) +
scale_x_log10() +
facet_wrap(type~unit)

```

## `stat\_bin()` using `bins = 30` . Pick better value with `binwidth` .



```
summary(subset(ISCN3$sample, grepl('bs|bs_sum', measurementID) & (value <= 0)))
```

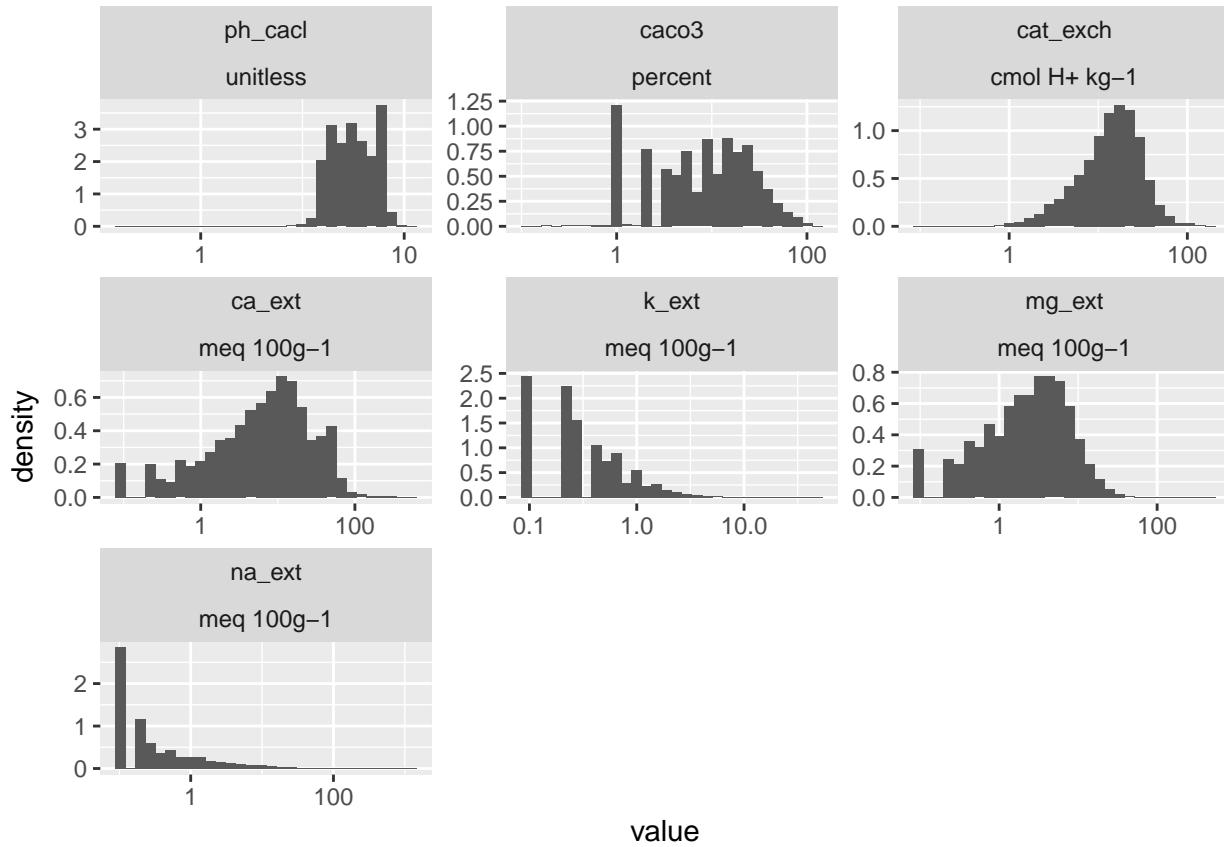
##	fieldID	measurementID	value	unit
##	00P07181:	2	bs_sum_01 : 1609	Min. :0 percent :2212
##	00P07182:	2	bs_01 : 603	1st Qu.:0 % : 0
##	00P07831:	2	13c_01 : 0	Median :0 cmol H+ kg-1: 0
##	01N05842:	2	14c_01 : 0	Mean :0 g cm-2 : 0
##	01N05843:	2	15n_01 : 0	3rd Qu.:0 g cm-3 : 0
##	01N05844:	2	al_dith_01: 0	Max. :0 meq 100g-1 : 0
##	(Other) :2200		(Other) : 0	(Other) : 0

## Extractable bases

- *ca\_al* TODO Luke N, what is this??
- *ca\_ext* calcium extractable (SSL: NH<sub>4</sub>OAc extractable)
- *k\_ext* potassium extractable (SSL: NH<sub>4</sub>OAc extractable)
- *mg\_ext* magnesium extractable (SSL: NH<sub>4</sub>OAc extractable)
- *na\_ext* sodium extractable (SSL: NH<sub>4</sub>OAc extractable)
- *base\_sum* sum of NH<sub>4</sub>OAc extractable bases

```
ggplot(merge(subset(ISCN3$sample, grepl('(ca|k|mg|na)', measurementID) & value > 0),
            ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density..)) +
  scale_x_log10() +
  facet_wrap(type~unit, scales='free')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
summary(subset(ISCN3$sample, grep('ca|k|mg|na', measurementID) & (value <= 0)))
```

```
##      fieldID      measurementID      value
## 05N04528:    5  na_ext_01 :87848  Min.   :-7.0e-01
## 06N03497:    5  k_ext_01  :27763  1st Qu.: 0.0e+00
## 07N01837:    5  caco3_01 :25036 Median  : 0.0e+00
## 07N01843:    5  mg_ext_01 :12471 Mean    :-4.9e-06
## 08N00509:    5  ca_ext_01 : 8912  3rd Qu.: 0.0e+00
## 08N01184:    5  ph_cacl_01:     2 Max.    : 0.0e+00
## (Other) :162003 (Other)    :     1
##                  unit
##  meq 100g-1  :136994
##  percent     : 25036
##  unitless    :     2
##  cmol H+ kg-1:     1
##  %           :     0
##  g cm-2      :     0
##  (Other)     :     0
```

## Metal (Al, Fe, Mn) extracts

For ISCN3 this is only for NRCS data.

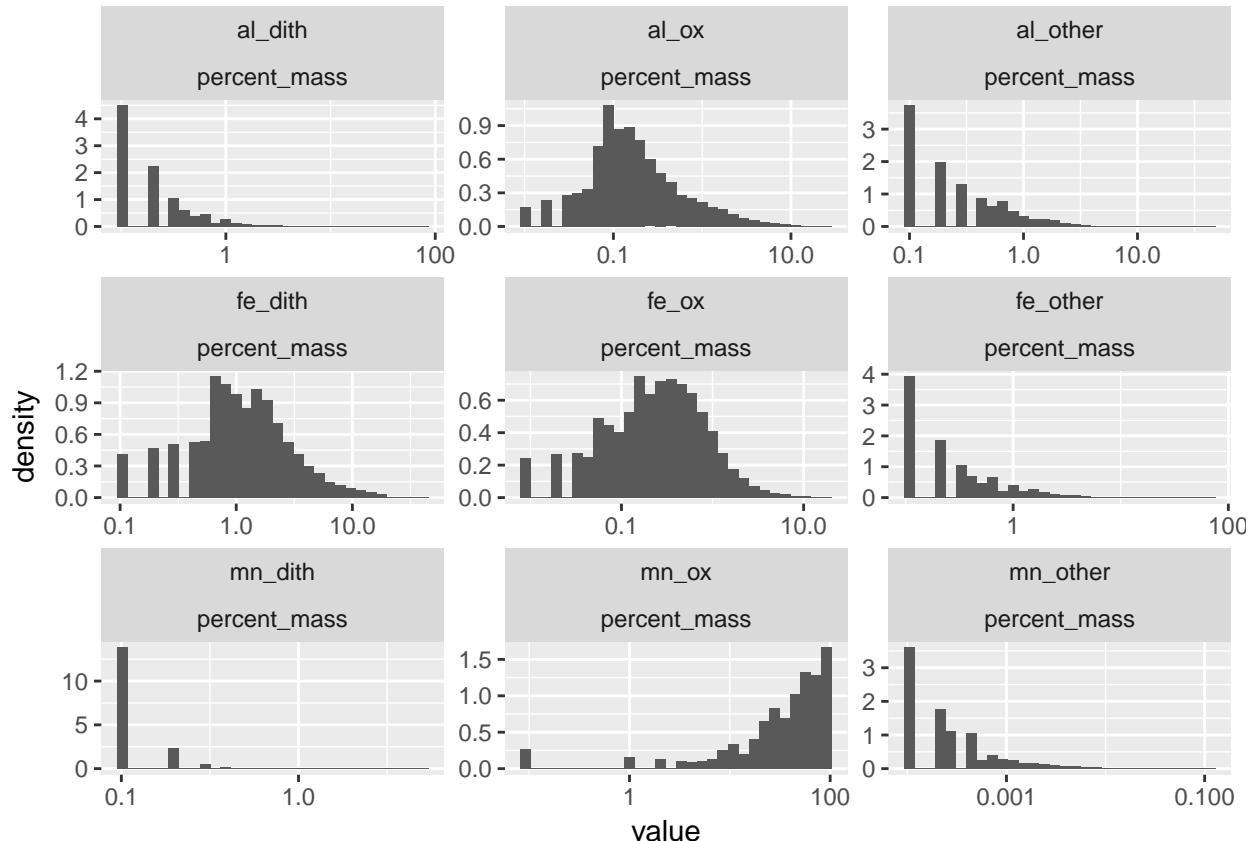
- \*<sub>dith</sub>\* Dithionite citrate extractable
- \*<sub>ox</sub>\* Oxalate extractable
- *al\_other* and *fe\_other* Sodium pyrophosphate extractable
- *mn\_other* KCl extractable

```

ggplot(merge(subset(ISCN3$sample,
                  grep('^(al|fe|mn)_', measurementID) &
                  (value <= 100 & value >= 0)),
           ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density..)) +
  scale_x_log10() +
  facet_wrap(type~unit, scales='free')

```

## Warning: Transformation introduced infinite values in continuous x-axis  
 ## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
 ## Warning: Removed 101203 rows containing non-finite values (stat\_bin).



```

summary(subset(ISCN3$sample, grep('^(al|fe|mn)_', measurementID) &
               !grep('ext', measurementID) & (value < 0 | value > 100)))

```

	fieldID	measurementID	value
##	95Z01392:	3 mn_ox_01 :27180	Min. : 100.1
##	95Z01393:	3 al_ox_01 : 14	1st Qu.: 209.5
##	95Z01395:	3 fe_ox_01 : 14	Median : 372.7
##	95Z01396:	3 fe_dith_01: 2	Mean : 590.1
##	95Z01397:	3 13c_01 : 0	3rd Qu.: 678.4
##	95Z01398:	3 14c_01 : 0	Max. : 35481.6
##	(Other) :27192	(Other) : 0	
##		unit	
##		percent_mass:27210	
##		% : 0	

```

##  cmol H+ kg-1:      0
##  g cm-2       :      0
##  g cm-3       :      0
##  meq 100g-1   :      0
##  (Other)      :      0

```

## Phosphorus

- $p_{bray}$  bray-1 extractable
- $p_{ox}$  oxalate extractable
- $p_{other}$  mehlich3 extractable

```

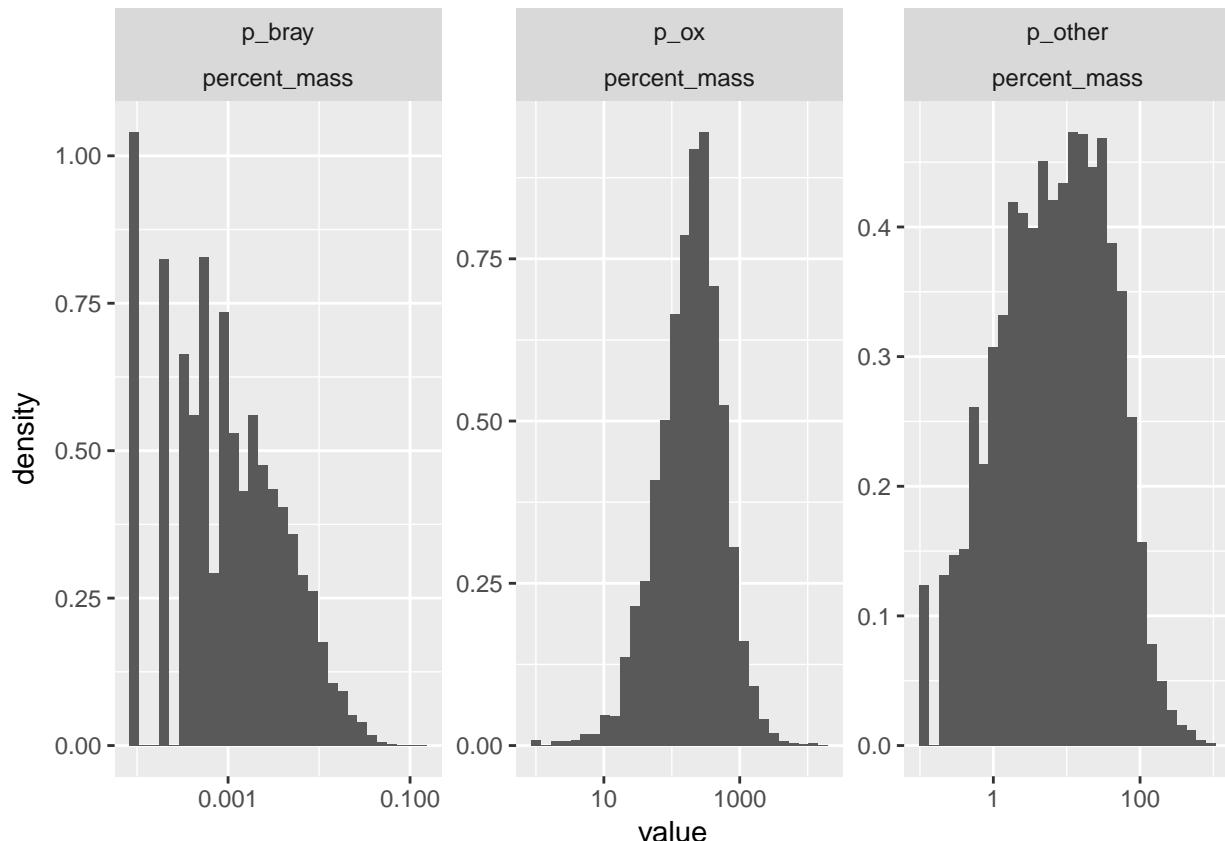
ggplot(merge(subset(ISCN3$sample, grepl('^p_',
ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density...)) +
  scale_x_log10() +
  facet_wrap(type~unit, scales='free')

```

```

## Warning in self$trans$transform(x): NaNs produced
## Warning: Transformation introduced infinite values in continuous x-axis
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 7887 rows containing non-finite values (stat_bin).

```



```

summary(subset(ISCN3$sample, grepl('^p_',
measurementID) & (value > 100 | value < 0)))

```

##	fieldID	measurementID	value
----	---------	---------------	-------

```

## 02N00030:    2 p_ox_01    :18254   Min.   : -0.2
## 02N00031:    2 p_other_01: 290    1st Qu.: 176.8
## 02N01473:    2 13c_01     : 0     Median : 277.4
## 02N03345:    2 14c_01     : 0     Mean    : 418.2
## 02N03354:    2 15n_01     : 0     3rd Qu.: 468.6
## 03N02430:    2 al_dith_01: 0     Max.   :16926.4
## (Other) :18532 (Other)   : 0
##               unit
## percent_mass:18544
## %           : 0
## cmol H+ kg-1: 0
## g cm-2       : 0
## g cm-3       : 0
## meq 100g-1   : 0
## (Other)      : 0

```

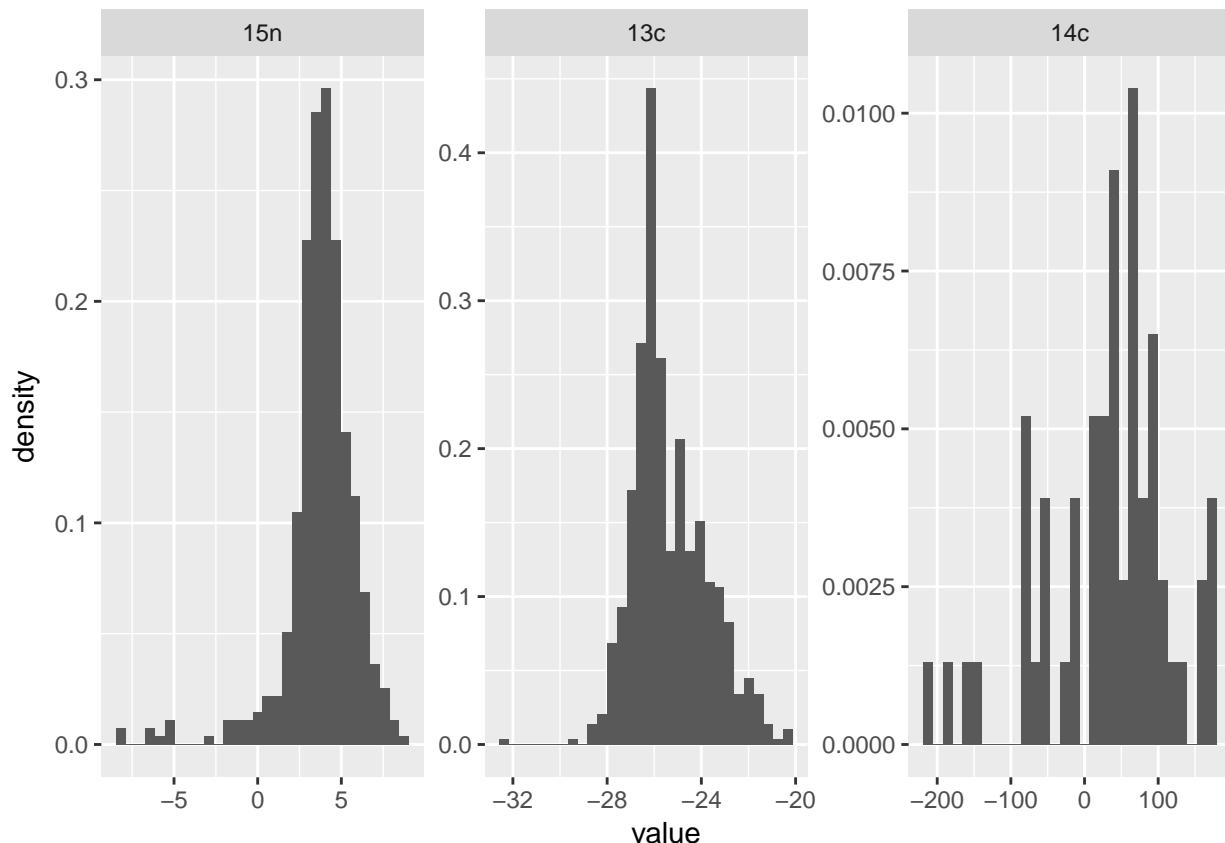
## Isotope

```

#15n          13c          14c
ggplot(merge(subset(ISCN3$sample, grep('1[345]', measurementID)),
            ISCN3$measurement)) +
  geom_histogram(aes(x=value, y=..density..)) +
  #scale_x_log10() +
  facet_wrap(~type, scales='free')

```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



```
summary(subset(ISCN3$sample, grep1('^1[345]', measurementID)))  
  
## fieldID measurementID value  
## BCPR 1.10 m,ab : 3 13c_01 :701 Min. :-208.90  
## BCPR 1.100 m,ab : 3 15n_01 :474 1st Qu.: -26.00  
## BCPR 1.20 m,ab : 3 14c_01 : 58 Median : -23.46  
## BCPR 1.40 m,ab : 3 al_dith_01 : 0 Mean : -11.53  
## BCPR 1.5 m, ab : 3 al_other_01: 0 3rd Qu.: 3.68  
## BCPR 1.60 m,ab : 3 al_ox_01 : 0 Max. : 175.90  
## (Other) :1215 (Other) : 0  
  
## unit  
## % :1233  
## cmol H+ kg-1: 0  
## g cm-2 : 0  
## g cm-3 : 0  
## meq 100g-1 : 0  
## percent : 0  
## (Other) : 0
```