

# GAINS - QAQC and Summary Statistics

## Status: Preliminary DRAFT

- Internal Use Only: Please to do not circulate or quote -

## Country Coverage

- All countries

## Data source

- Origin: Surveillance data extracts retrieved from <http://gains.org> in 7-day tranches
- Date Extracted: **27-Nov-12**

## Prepared by

- Preston, Murray, Zambrano-Torrel (EcoHealth Alliance NYC)

## Caveats

- This report focuses on surveillance samples (rows) in GAINS
- The number of samples taken per animal varies widely with a mean of 5.3
- These results are intended to guide QAQC of the GAINS database for modeling
- This is not a progress report on Predict
- **Data entry and validation efforts are ongoing**
- This report is automated and **formatting may vary**.

## Intended Recipients:

- **WCS Nanaimo:** Damien Joly, Tammie O'Rourke, Sarah Olson
- **UC Davis:** Christine Kreuder Johnson

## Data Wrangling

- Total number of surveillance records

```
GAINS: n = 173960
```

- Remove parentheses, asterisks, and spaces from column names:

```
Out[249]: Index([SiteName, GAINS3_EventName, Country, StateProv, District, Latitude, Longitude, HabitatType, HabitatComment,
LandscapeConversionGradient, AnthropogenicChange, AnthropogenicChangeSecondary, DomesticAnimals, OtherDomesticAnimals,
DateStarted, TimeStart, TimeDuration, RecorderID, RecorderOther, RecorderAffiliation, ArchivedData, SourceCertainty,
PrimaryInterface, OtherPrimaryInterface, SecondaryInterface, OtherSecondaryInterface, ReasonForCollection, KnownSource,
EventCustomColumns, ContainerID, ContainerType, Quantity, QuantityCertainty, QuantityUnit, SampleUnitID, AnimalIDGAINS,
AnimalIDRecordedAs, SampleDate, sampleUnitLatitude, sampleUnitLongitude, IdentifiedBy, CircumstanceObservation, SampleLoc
AnimalClassification, AnimalTaxa, TaxonomicDescriptor, SpeciesScientificName, SpeciesCommonNameEnglish, SpeciesCommonName
IdentificationCertainty, TagBandID, Sex, AgeClass, Pregnant, WithYoung, Lactating, ConditionAtSampling,
ReportedAsConditionAtCapture, ConditionAtRelease, PreservationMethod, OtherConditionPreservation, Use, OtherUse,
DestinationOfProduct, OriginKnown, OriginLocation, TimeSinceOrigin, Euthanasia, ClinicalSignsGrossFindings, DiseaseSuspec
SuspectedDisease, NecropsyExam, FieldNecropsyResults, BodyMeasurementBodyPartType, BodyMeasurementUnit, BodyMeasurementVa
Part, PartDescription, PerceivedRateOfSale, PerceivedRateSaleTimeUnit, PriceUnitLocal, PriceUnitUSD, VendorQuantity,
VendorQuantityUnit, VendorQuantityCertainty, ButcheringActivity, Butcher, ButcherObservableRiskFactors,
ButcherReducesBloodContact, Quarter, sampleUnitNotes, sampleUnitCustomColumns, SpecimenID, SpecimenIDBarcode, SpecimenDat
SpecimenType, SpecimenTypeDescriptor, SpecimenCondition, SpecimenComments, SamplingMethod, SampleAliquoted, AliquotID,
SampleContainer, StorageMedium, FieldStorageMethod, LaboratoryStorageMethod, SpecimenLocationFacilityName,
SpecimenLocationInLab, SpecimenLocationFacilityAddress, TimeFromCollectionToFreezing, BloodCollectionTube, PlasmaCollecte
SerumCollected, MethodOfSerumPlasmaCollection, VolumeOfSerumPlasma, Notes, SpecimenCustomColumns, DateCreated, EntryID,
InPurgatory], dtype=object)
```

- Remove Pre-FY1 data

```
Post-FY1: n = 145181
```

- Number of samples in purgatory

```
Out[251]: 3437
```

- Percent samples in purgatory

```
2.3673896722 %
```

- Remove samples in purgatory

```
n = 141744
```

- Rename organizations with consistent acronyms

```
Out[254]: array([WCS, UC Davis, EHA, GVFI, Smithsonian], dtype=object)
```

- Standardize missing and unknown values

- Create dataframe with core columns

```
Out[255]: Index([SiteName, Country, StateProv, Latitude, Longitude, AnthropogenicChange, DomesticAnimals, DateStarted, RecorderID, RecorderAffiliation, PrimaryInterface, ReasonForCollection, AnimalIDGAINS, AnimalIDRecordedAs, IdentifiedBy, AnimalClassification, TaxonomicDescriptor, SpeciesScientificName, IdentificationCertainty], dtype=object)
```

## Summaries

### Summaries of key fields (columns) for sample records in GAINS (rows)

```
Out[256]:
```

	SiteName	Country	StateProv
count	141744	141744	141744
unique	893	20	191
top	Libonga	Bangladesh	Dhaka
freq	4517	22998	10167

```
Out[257]:
```

	Latitude	Longitude
count	141505.000000	141505.000000
mean	51.295507	155.126959
std	5063.469931	14149.684793
min	-90.294580	-99.469722
25%	-3.125900	-48.083394
50%	3.625400	18.642920
75%	18.116220	90.040500
max	549872.000000	1536601.000000

```
Out[258]: <class 'pandas.core.frame.DataFrame'>
Index: 4 entries, count to freq
Data columns:
AnthropogenicChange    4    non-null values
DomesticAnimals         4    non-null values
DateStarted             4    non-null values
RecorderID              4    non-null values
RecorderAffiliation     4    non-null values
dtypes: object(5)
```

```
Out[259]: <class 'pandas.core.frame.DataFrame'>
Index: 4 entries, count to freq
Data columns:
PrimaryInterface        4    non-null values
ReasonForCollection     4    non-null values
AnimalIDGAINS           4    non-null values
AnimalIDRecordedAs      4    non-null values
IdentifiedBy            4    non-null values
AnimalClassification    4    non-null values
dtypes: object(6)
```

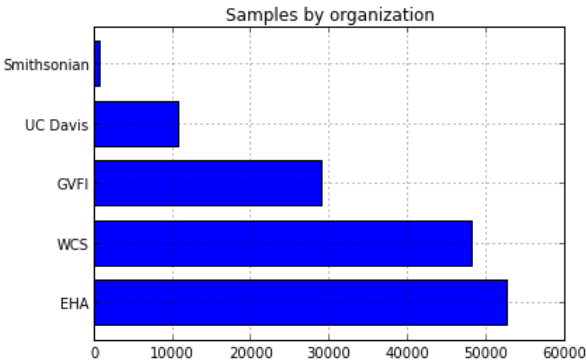
Out[260]:

	TaxonomicDescriptor	SpeciesScientificName	IdentificationCertainty
count	118624	133159	140926
unique	1038	1113	3
top	Macaca fascicularis	Pteropus giganteus	Actual
freq	6862	8305	115175

Sample Counts

Out[262]: EHA 52722  
WCS 48281  
GVFI 29158  
UC Davis 10767  
Smithsonian 816

Out[263]: <matplotlib.axes.AxesSubplot at 0x1201be450>



Counts by Country

Out[264]: Bangladesh 22998  
Cameroon 20075  
Bolivia 16939  
Malaysia 11287  
Lao Peoples Democratic Republic 9544  
Peru 8329  
Mexico 7961  
Congo, Republic Of 7878  
Congo, Democratic Republic Of 7025  
Brazil 6534  
Cambodia 4882  
Tanzania, United Republic Of 4819  
Uganda 3215  
Gabon 3032  
Rwanda 2384  
China 2052  
Vietnam 1016  
Thailand 973  
Colombia 407  
India 394

Site name (SiteName)

```

Out[265]: Libonga 4517
          Houay Dturn Village 2392
          Shuvarampur, Amicapur 2356
          San Luis Chico 2330
          Master colony, Faridpur 2314
          Limbe Wildlife Centre, Limbe Cameroon 2261
          Peroto - Evento J1 2115
          Ambicapur, Faridpur 1808
          ENB-0001-0137 1781
          Okoroba 1773
          Selangor 1662
          Kole 1613
          Sara Ana 1573
          Srimongol 1556
          Kasi Market 1510
          ...
          Bakoumba 23 1
          Bakoumba 22 1
          Bakoumba 20 1
          Bakoumba 2 1
          Bakoumba 19 1
          Bakoumba 18 1
          Bakoumba 17 1
          Bakoumba 16 1
          Bakoumba 15 1
          Bakoumba 14 1
          Bakoumba 13 1
          Bakoumba 12 1
          Bakoumba 11 1
          Bakoumba 10 1
          Avangane, about 40km east of Nkoteng 1
          Length: 893

```

## State/Province (StateProv)

```

Out[266]: Dhaka 10167
          Southwest 7504
          Sangha 5967
          Centre 5357
          Selangor 4213
          Vientiane 3883
          Iringa 3643
          Chiapas 3492
          Campeche 3366
          South 3348
          Faridpur 3100
          Perak 2912
          Sylhet 2426
          Houaphan 2392
          Kasai Oriental 2363
          Ballivián 2322
          Kasai Oriental 2318
          Abel Iturralde 2190
          San Andres segunda sección 2115
          Chittagong 1936
          East Africa 1926
          Amazonas 1798
          Nor Yungas 1701
          Kandal 1692
          Loreto 1633
          Ucayali 1480
          Faridpur 1476
          Johor 1459
          Sud Yungas 1416
          Ballivian 1402
          Lekoumou 1328
          Roraima 1312
          East 1241
          Yolosa 1207
          Murillo 1199
          Battambang 1199
          Lima 1171
          Haut-Ogooué 1126
          Marban 1037
          Manaus 965
          Rangpur 910
          North 910
          Maranhão 874
          Tangail 828
          Nyanga 816
          Cercado 804
          Velasco-Alfaro de Chavez 730
          BTB 726
          Yunnan 723

```

Westren Uganda	723
Haut Ogooue	720
Rajshahi	713
Saravan	688
Northern Province	687
Maynas / Loreto	648
Littoral	633
EASTERN PROVINCE	621
Khammouane	608
Lam Dong Province	600
Xieng Khouang	585
Bambama	563
Tumbes	562
Ucayali	550
Pulau Pinang	548
Distrito Federal	546
Huancabamba-Piura	540
Bolikhamxay	534
Negeri Sembilan	524
Pahang	499
Iringa	490
Northwest	475
Adamaoua	466
Madre de Dios	465
Rajbari	458
Kinshasa	451
Minas Gerais	445
Kinshasa Province	440
KAD	440
Comilla	432
Kinshasa	424
Campeche	418
Cundinamarca	407
Province Oriental	405
Champasak	402
Goiás	400
Tamil Nadu	394
Hainan	390
NORTHERN PROVINCE	386
Mbeya	386
Buhoma	368
Manaus/Amazonas	360
Ogooué Ivindo	340
Srakaew	335
Guang Xi	334
North Province	321
Lam Dong	320
Negeri Sembilan	278
La Paz	264
Chachoengsao	256
Ascension de Guarayos	250
North Kivu	240
Jaen	234
Bascongo	227
Selangor	224
San Martin	219
Puno	217
Lalmonirhat	212
Preah Vihear	210
centre	199
Shanxi	196
Guangdong	191
Presidente Figueiredo/AMAZONAS	176
Barishal	176
southwest	169
Cusco	169
Champassak	166
Sylhet	164
Unknown	158
Caranavi	157
Sekong	152
Kuala Lumpur	149
Santa Cruz - Jaén	144
Presidente Figueiredo/Amazonas	144
Iringa	134
Sabah	133
Ratanakiri	133
Johor	132
GuangXi	132
Kampong Chhnang	130
Chantaburi	129
Estado de México	124
Nakhon Ratchasima	120
Western Province	109
Quispicanchi / Cusco	102
Trat	93
Ruvuma	84

Iringa	82
Villa Tunari	75
KIGALI TOWN	73
PHP	72
Savannakate	64
Stung Treng	61
Rio Grande do Sul	60
Ucayali / Coronel Portillo	54
Attapue	54
South Kivu	52
Siem Reap	51
Selangor and Johor	50
Kinshasa	48
Chiclayo	47
Jiangxi	46
Quillacollo	45
PIURA	44
Vinh Phuc province	42
Huanuco	40
Ratanakiri	39
Vinh Phuc Province	38
South Western Uganda	36
South Kivu and Maniena Provinces	33
Kampong Cham	32
Pursat	31
Northern	31
Haut Ogoou	30
Ratanakiri province	28
Pang-nga	28
Cochabamba	25
North Kivu Province	24
Preah Vihear Station	23
WESTERN PROVINCE	18
Sichuan	18
Western Region	16
Attapeu	16
campeche	15
Province du Nord	15
Pool	15
Mondulkiri	15
Kanungu	14
East Affrica	11
Piura	10
Ninh Binh	10
Bunyoro	10
Anhui	10
Melaka	8
Hunan	8
Phang Nga	7
Hanoi	6
Abim	6
Sakaeo	5
Cuvette	5
Guizhou	4
soutwest	1
Length: 191	

## Recorder ID (RecorderID)

Out[267]:	arif@ecohealthalliance.org	10048	
	Lee Chee Yen	7615	
	Patricia Mendoza	6856	
	Kongsy KHAMMAVONG	6262	
	Oscar Rico	5760	
	Ahmadou	5274	
	Joseph Diffo	4550	
	Jean Michel	4455	
	Rosario Barradas	3686	
	Mei-Ho Lee	3672	
	Jean Michel Takuo	3460	
	BOUNGA Gerard	3441	
	Maria Makuwa	3197	
	Erika Alandia	2794	
	Dr.Ariful islam	2314	
	Catia Dejuste de Paula	2215	
	Kenneth Cameron	2129	
	Rolando Limachi Qui	2115	
	Priscilla Joyner	1784	
	Abdulhameed Kateregga	1771	
	Dr.Ariful Islam	1556	
	Monica Romero Solorio	1495	
	Dr. Najmul Haider, Dr. Ausraful Islam, Dr. Ariful Islam, Dr. Suman Kumar Paul	1476	
	Kirsten Gilardi	1465	
	Oswaldo Palabral	1416	
	grayala@wcs.org	1409	

Soubanh Silithammavong	1334	
Dr.Kevin J Olival	1319	
Maria MAKUWA	1222	
Aristide	1182	
Isabel Moya	1162	
Jesus Sotomayor	1133	
Mathieu Bougarel	1126	
Dr Julius NZIZA	1102	
Zikankuba Sijali	1072	
Jose Luis Mollericon	1028	
Hap Sovath	987	
Supaporn Wacharapluesadee	933	
John Calder Escalante	918	
Dr. Ariful Islam, Dr. Sumon Kumar Paul	910	
Chor Kimashalen	866	
Joseph Atibu Losoma	818	
Chris Whittier	816	
Ms Phinouluck (Students)	815	
Ken Cameron	795	
mbarae@gmail.com	765	
virginie.rougeron@gvfi.org	720	
Dr. Asad	713	
Fabiola Suarez	710	
Dr. Salah uddin Khan, Dr. Najmul Haider, Dr. Atiqul islam, Dr. Suman Kuamr Paul	698	
Rafael Avila	670	
lclark@wcs.org	654	
Dr. Najmul Haider, Dr. Golam yasin.	650	
Maria Makuwa/ Joseph Atibu Losoma	641	
Marina Seixas	640	
Dr. Ariful Islam	621	
Guangjian Zhu	614	
Flavia Miranda	595	
Bounga Gerard	591	
Luciana Mendiola	539	
nava@ecohealthalliance.org	510	
GuangJian Zhu	506	
Harrison Sadiki (hsadily@gmail.com)	504	
katereggaa@gmail.com	478	
Glenda Ayala	478	
Ausraful Islam	458	
djwolk@gmail.com	435	
Mwokozi Mwanzalila	434	
Rolando Limachi	431	
Victoria Pereira	407	
Nguyen Thi Thanh Nga	404	
Dr. Ariful islam	400	
Lucy KEATTS	397	
Adam Konecny	374	
CIRMF	370	
Alessandra Nava	351	
Mat LeBreton	345	
Dr. Najmulh haider	341	
Renata hurtado	335	
Zhu Guangjian	334	
Jim Desmond, Ausraful Islam,	320	
Zikankuba Sijali (zikankubasijali@gmail.com)	313	
Joseph Dikko Ledoux	313	
Md. Asadur Rahman, Sumon Kumar Paul	312	
Lucy Keatts	290	
Muhiddin Hatibu Salehe- mmkecha@gmail.com	288	
Lucy Ogg Keatts	288	
JOSEPH ATIBU LOSOMA	281	
Ivana Ghione	274	
Zikankubasijali@gmail.com	273	
Shalen	273	
Ahmadou Nana	273	
Wivine Mouellet	259	
Vannasin INOUDOM (Student)	245	
Mwokozi Mwazalila	244	
Jesus Sotomayot	243	
Zenna Babu (zmchomvu@gmail.com)	236	
Renata Hurtado	234	
Daniel Montes	229	
mmkecha@gmail.com	228	
JMT	223	
Dr H Raghuram	222	
Dr. Salauddin khan	212	
chenjp@gdei.gd.cn	211	
Pedro Mayor (mayorpedro@hotmail.com)	209	
Junpeng Zhang	196	
Herminio Ticona	195	
jin ping chen	191	
chor kimashalen	180	
zikankubasijali@gmail.com, hsadily@gmail.com	176	
Ausraful Islam	176	
Md. Asadur Rahman, Sumon Kumar Paul	164	
julius nziza , nzizavet@gmail.com	164	

Katherine Claros	162
Glenda Ayala	157
Zikankuba Sijali@gmail.com	150
Cameron	145
Jesus Sotmayor	140
zikankubasijali@gmail.com, hsadily@gmail.com, mmkecha@gmail.com	138
julius Nziza /nzizavet@gmail.com	116
Joseph ATIBU LOSOMA	103
JLD	101
Adriana Orellana	101
zikankubasijali@gmail.com	100
Jim Desmond	100
Angela Yang	100
Aizar Terrazas	99
Zikankuba Sijali/Zenna Babu	98
Dr. Ausraful Islam, Dr. Suman Kumar Paul, Dr. Atiqul Islam	98
Ausraful Islam	96
Soubanh SILITHAMMAVONG	92
Pedro Mayor	91
Dr. H. Raghuram	88
katchouang	87
Dr H Raghuram	84
Chor kimashalen	82
Shalen	81
Pasteur Institute	76
Matthew LeBreton	76
Zikankuba Sijali (zikznkubasijali@gmail.com)	73
Mat Lebreton	69
Heng cheavleng	68
Ivan Rodriguez	64
moiravet@hotmail.com	60
Sarah Olson	60
Mbabazi Racheal	60
Sample collector	57
Alain Ondzie	54
Missing	51
aristide	48
Monica Romero	46
Leanne Clark	42
EHA	40
Glenda Ayala (grayala@wcs.org)	36
Ivan Rodriguez, Kyle	30
mendiola@ecohealthalliance.org	28
Institute Pasteur Cambodia	28
Houmpheng Phomsy (Student)	27
Daniel Montes (dmontes@faunavet-peru.com)	26
Monica Romero Solorio	25
Maria Makuwa Joseph Atibu Losoma	25
Jacqueline KIMOBALA	25
Ivan Rodriguez	23
Rolando Limachi, Jose Luis Mollericon, Ivan Rodriguez	22
Kirsten	21
Dr. Suman Kumar Paul, Dr. Atiqul Islam	16
Chea Sokha	16
Oscar rico	15
Erika Alandia (ealandia@wcs.org)	8
AN	8
Mbabazi Racheal, Uganda Field Assistant	6
Akongo Marc-Joel	5
Length: 172	

## Anthropogenic Change (AnthropogenicChange)

Out[268]: Missing	46788
established village or indigenous settlement	24615
none of these	20053
area with hunting	14476
area with ecotourism	11291
ongoing deforestation	8493
crop growing	6726
urban	3539
ongoing road building	1900
area of new human settlement	1393
livestock grazing area	920
mineral mining	772
pristine undeveloped area	529
intensive livestock holding	152
crop growing and livestock grazing area	97

## Domestic Animals (DomesticAnimals)



```

Out[269]: Missing          47876
multiple domestic animal spp 35338
none                        25263
multiple livestock spp      14193
chickens/ducks              9409
dogs/cats                   4762
cattle                      1961
other                       1958
goats/sheep                 742
other (please describe)     242

```

## Primary Interface (PrimaryInterface)

```

Out[270]: free-ranging; contact with domestic animals or humans not likely 45725
hunted 28926
peri-domestic/in or near human dwelling(s) 18511
sanctuary 7096
for sale in large market (> 20 vendors) 5639
temporary holding facility 4656
other 4317
Market 3966
public safety hazard (e.g. threat to human) 3598
Freeranging 3147
for sale in medium market (5-20 vendors) 2483
zoo 2015
contact with tourists/ecotourism 1712
for sale in small market (< 5 vendors) 1664
for sale in restaurant 1531
raiding crops 1406
Wetmarkets 1083
rehabilitation center 939
protected area management 691
contact with park personnel/intensive wildlife management area 685
wild animal farm 536
domestic/near human dwelling(s) 324
preying on livestock or their food 312
Missing 288
private wildlife collection or pet 285
contact with workers in extractive industry 193
Confiscation 16

```

## Animal Classification (AnimalClassification)

```

Out[271]: Wild          110087
Captive Wild          30672
Domestic              616
Missing              334
Feral Domestic        35

```

## Taxonomic Descriptor (TaxonomicDescriptor)

```

Out[272]: Missing                23120
Macaca fascicularis             6862
Pteropus giganteus              4675
Pteropus                        3468
Cephalophus monticola          2358
Pteropus giganteus             2130
Mastomys                        1982
Pan troglodytes                 1971
Rattus rattus                   1866
Cynopterus                      1793
Tayassu pecari                  1716
Manis javanica                  1668
Rousettus sp.                   1427
Rousettus leschenaulti         1406
Lagothrix lagotricha           1249
...
Dendrohyrax sp.                 1
Cynopterus sphinx               1
Crossarchus sp.                 1
Chelus sp.                      1
Chaerephon plicata              1
Cercopithecus ascanius          1
Cercoipetecus wolfi            1
Cephalophus silvicultor         1
Carnivora sp. sp.              1
Bothrops sp.                   1
Aves Sp.                        1
Atherurus africanus centralis   1
Anatidae sp.                   1
Amphibia                       1
Agapornis roseicollis           1
Length: 1039

```

### Primary Interface (PrimaryInterface)

```

Out[273]: free-ranging; contact with domestic animals or humans not likely 45725
hunted 28926
peri-domestic/in or near human dwelling(s) 18511
sanctuary 7096
for sale in large market (> 20 vendors) 5639
temporary holding facility 4656
other 4317
Market 3966
public safety hazard (e.g. threat to human) 3598
Freeranging 3147
for sale in medium market (5-20 vendors) 2483
zoo 2015
contact with tourists/ecotourism 1712
for sale in small market (< 5 vendors) 1664
for sale in restaurant 1531
raiding crops 1406
Wetmarkets 1083
rehabilitation center 939
protected area management 691
contact with park personnel/intensive wildlife management area 685
wild animal farm 536
domestic/near human dwelling(s) 324
preying on livestock or their food 312
Missing 288
private wildlife collection or pet 285
contact with workers in extractive industry 193
Confiscation 16

```

### Species Scientific Name (SpeciesScientificName)

```

Out[274]: Missing      8585
          Pteropus giganteus      8305
          Macaca fascicularis      6865
          Cephalophus monticola    3969
          Rattus rattus             2764
          Calomys sp.              2381
          Carollia perspicillata    2187
          Pteropus giganteus        2130
          Pan troglodytes           2086
          Artibeus lituratus        1892
          Tayassu pecari            1718
          Manis javanica            1668
          Suncus murinus            1657
          Rousettus leschenaulti    1526
          Cephalophus callipygus    1281
          ...
          Euoticus pallidus         1
          Epixerus ebii             1
          Cynopterus sphinx         1
          Chelus fimbriatus         1
          Cercopithecus ascanius    1
          Cercopithecidae           1
          Carnivora sp.             1
          Canis adustus             1
          Buteo sp.                 1
          Bothrops sp.             1
          Black Crested Mangabey Lophocebus aterimus 1
          Bitis nasicornis          1
          Bitis gabonica            1
          Amphibia                  1
          Agapornis roseicollis     1
          Length: 1114

```

#### Identification Certainty (IdentificationCertainty)

```

Out[275]: Actual      115175
          Estimate     25130
          Missing       818
          Unknown       621

```

#### Field Storage Method (FieldStorageMethod)

```

Out[276]: Liquid Nitrogen      49422
          Missing               41773
          Ambient temperature    15771
          Ice cooler             11097
          Freezer (-80C)         6000
          Dry shipper            3693
          Envelope with Dessicant 3423
          Silica Dessicant       3030
          Freezer (-20C)         2548
          Clip lock bag          1900
          N/A                    1158
          Dry ice                 683
          Sealed Container       599
          Unsealed container (ambient temperature) 342
          Warm: >25-30oC         305

```

#### Laboratory Storage Method (LaboratoryStorageMethod)

```

Out[277]: Freezer (-80C)       70734
          Missing               46592
          Ambient temperature    11924
          Liquid Nitrogen        2923
          Envelope with Dessicant 2556
          Freezer (-20C)         2451
          N/A                    1420
          Sealed Container       894
          Clip lock bag          789
          Ice cooler             480
          Unsealed container (ambient temperature) 268
          Cool: approximately 4oC 264
          Room temperature: 20-25oC 157
          Silica Dessicant       105
          Dry ice                 104
          Freezer (-85C)         83

```

#### Specimen Location Facility Name (SpecimenLocationFacilityName)

Out[278]: Missing	41183
ICDDR,B Lab	9460
National Animal Health Centre	9274
NAMRU-6, Lima	4552
FMVZ, UNAM	4381
Instituto de Biología Molecular y Biotecnología (IBMB) - UMSA	3630
WCS-Congo	3381
Sokoine University of Agriculture(SUA)	3357
UMSA	3020
ICDDR,B lab	2938
FMVZ	2764
ICDDR,B Lab.	2536
Instituto de Biología Molecular y Biotecnología (IBMB) - UMSA	2512
INRB/GVFI Kinshasa	2360
Instituto de Biología Molecular y Biotecnología (IBMB)-UMSA	2333
ICDDR,B Lab	2314
WCS Store Room	2248
ICB/USP	1939
BSL-2,Animal lab	1781
ICDDR,B Lab and CII Lab,USA	1760
IBMB - UMSA	1547
WCS Manaus	1498
BSL3 Virology laboratory ICB II	1384
ICDDR,B Lab	1319
FMVZ - USP	1312
WDDL	1308
WARM DEPT	1264
Pasteur Institute of Cambodia	1238
WCS-Bolivia	1220
PERHILITAN HQ, Ex-situ Conservation Department, Treatment room	1070
WHO-cc Chulalongkorn	933
INS, LIMA	918
WARM Department	839
Oficina WCS - La Paz, Bolivia	775
Wuhan Institute of Virology, Chinese Academy of Sciences	754
INRB-Predict Lab	751
GVFI Cameroon lab	733
CIRMF	720
INRB-PREDICT Lab, Kinshasa	696
FMVZ UNAM	680
IBMB	612
WCS store room	595
WCS office	549
WCS BZV	534
WARM DEPT.	506
BSL-2, Animal lab	458
Iringa & SUA Morogoro	452
Wildlife Conservation Society	447
NAMRU 6, Lima	444
BSL-2, Animal Lab	432
HUA veterinary Microbiology and Infectious disease laboratory	416
LAB_NAMRU-6, Lima	396
Chinese Academy of Sciences,Wu han Institute of Virology	390
Yunnan Institute for Endemic Diseases control and Prevention	374
Lam Dong FPD	342
N/A	317
INS	301
SUA Morogoro	288
Morogoro (SUA)	282
Rwanda	280
HUA Vet Lab	258
Freezer (-80C)	246
MGVP Musanze Lab	226
NAMRU	218
refreezer	211
NAMRU 6 , Lima	194
Lab of Jinping Chen	191
WCS, Lima	185
MGVP Musanze	174
Laboratorio de Diagnóstico Veterinario	157
Iringa (HALI)	153
NAMRU 6, Iquitos	150
Instituto de Biología Molecular y Biotecnología (IBMB)-UMSA	149
Instituto de Biología Molecular	139
Sabah Wildlife Rescue Centre Laboratory	133
East China Normal University	132
National Animal Health Center	127
INRB-Kinshasa -DRC	122
EIP_NAMRU-6, Lima	110
FUSM	106
Institut Pasteur Phnom Penh	104
WCS - Bolivia	98
Carrera de Biología-Universidad Mayor de San Andrés	97
CES, IISc, Bangalore	88
WCS- Bolivia	76

Instituto de Biología Molecular y Biotecnología	76
Instituto de Biología Molecular y Biotecnología	76
MGVP Goma	66
MGVP Baltimore	54
LAB_NAMRU-6	54
NAMRU-6, Iquitos	53
UPCH, Lima	44
Cool: approximately 4oC	44
MGVP Lab	40
CS Pachitea, Piura	40
WCS - Congo	29
MGVP Musanze, Rwanda	28
WCS-BOLIVIA	22
MGVP Rwanda	21
WCS Bolivia	18
WCS, Iquitos	15
Wildlife Conservation Society Lao Program	10
MGVP DRC (Lwiro)	9
MGVP DRC	9
EIP_NAMRU-6	9
BZV	9
2.4	9
MGVP DRC and US	8
Faculty of Medicine	7
WCS - Bolivia, La Paz	6
Museo Nol Kempff Mercado	5
MGVP Biobank	5
Diresa, Tumbes	5
Office, BZV	4
MGVP- Musanze Lab	4
iNRB-Predict Lab	3
Wildlife Conservation Society Lao PDR	3
WCS Laos	3
MGVP Buhoma Lab	2
LAP_UNMSM, Lima	2
LAP_UNMSM	2
Instituto de Biología Molecular-IE/UMSA	2
INS_Reference Lab, Pucallpa	2
Anubis Vet_Lima	2
ZDDL, BZV	1
MGVP UC Davis	1
MGVP DRC (Lwiro)	1
Length: 127	

## Plasma Collected (PlasmaCollected)

```
Out[279]: Missing      88042
          N/A          26940
          No           25042
          Yes           1720
```

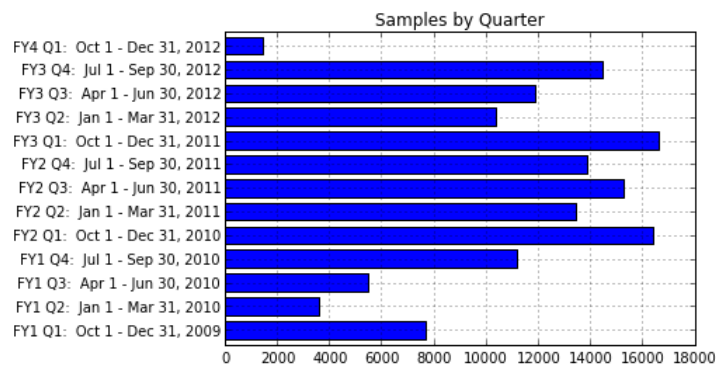
## Serum Collected (SerumCollected)

```
Out[280]: Missing      90128
          N/A          26796
          Yes          15029
          No           9791
```

## Quarter

```
Out[281]: FY1 Q1: Oct 1 - Dec 31, 2009      7695
          FY1 Q2: Jan 1 - Mar 31, 2010       3632
          FY1 Q3: Apr 1 - Jun 30, 2010       5468
          FY1 Q4: Jul 1 - Sep 30, 2010      11184
          FY2 Q1: Oct 1 - Dec 31, 2010      16401
          FY2 Q2: Jan 1 - Mar 31, 2011      13468
          FY2 Q3: Apr 1 - Jun 30, 2011      15256
          FY2 Q4: Jul 1 - Sep 30, 2011      13878
          FY3 Q1: Oct 1 - Dec 31, 2011      16606
          FY3 Q2: Jan 1 - Mar 31, 2012      10362
          FY3 Q3: Apr 1 - Jun 30, 2012      11885
          FY3 Q4: Jul 1 - Sep 30, 2012      14470
          FY4 Q1: Oct 1 - Dec 31, 2012       1439
```

Out[282]: <matplotlib.axes.AxesSubplot at 0x10ab66050>



Origin Known (OriginKnown)

Out[283]:

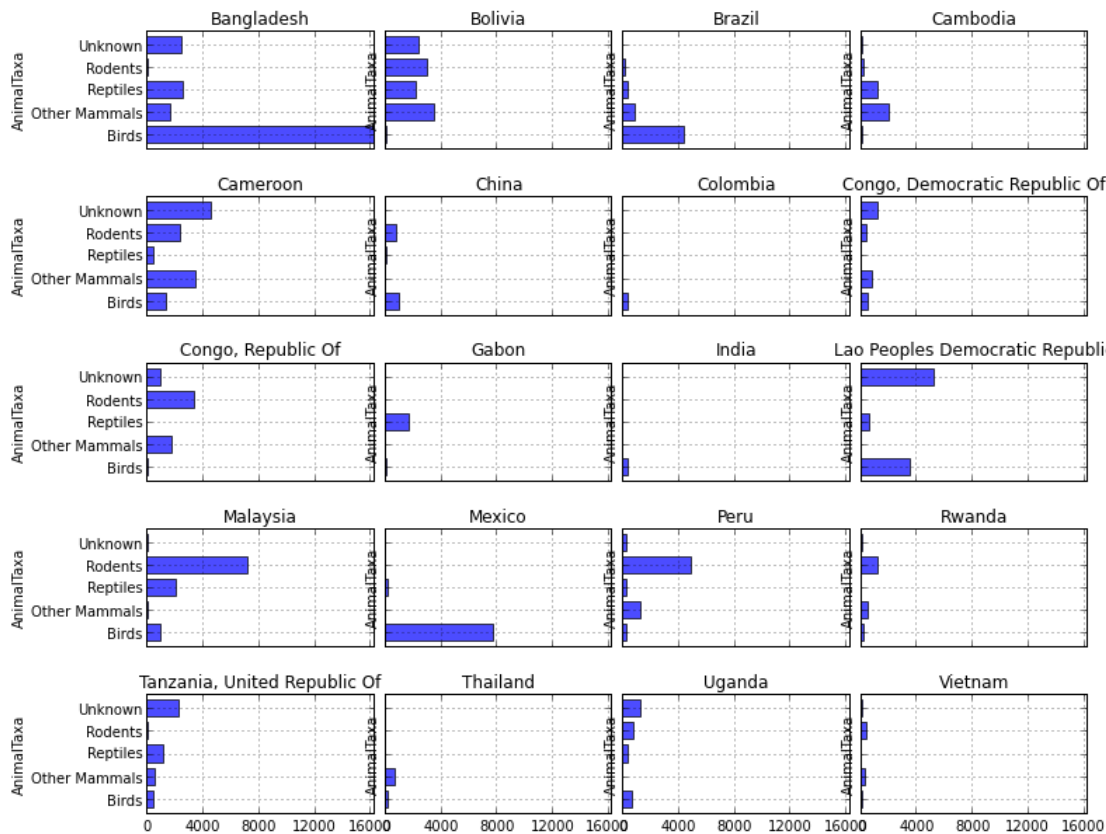
Missing	116083
Known	13633
Unknown	7882
Unknown but local	4146

Pregnant

Out[284]:

Missing	61605
No	45217
N/A	20618
Unknown	7425
Yes	6502
False	377

Sample counts by country and taxa



Out[289]: Country	AnimalTaxa	
Bangladesh	Bats	16229
	Birds	1642
	Other Mammals	2542
	Primates	60
	Rodents	2525
Bolivia	* Not Yet Typed	102
	Bats	3546
	Birds	2269
	Other Mammals	3017
	Primates	2412
	Reptiles	772
	Rodents	4821
Brazil	Bats	4448
	Birds	892
	Other Mammals	383
	Primates	232
	Reptiles	4
	Rodents	575
Cambodia	* Not Yet Typed	101
	Bats	2086
	Birds	1238
	Other Mammals	208
	Primates	108
	Reptiles	6
	Rodents	1132
	Unknown	3
Cameroon	* Not Yet Typed	1411
	Bats	3500
	Birds	462
	Other Mammals	2367
	Primates	4632
	Reptiles	115
	Rodents	1924
	Unknown	1921
China	Bats	1021
	Birds	48
	Other Mammals	143
	Rodents	840
Colombia	Birds	353
	Other Mammals	15
	Primates	33
	Rodents	6
Congo, Democratic Republic Of	* Not Yet Typed	512
	Bats	873
	Birds	17
	Other Mammals	444
	Primates	1244
	Reptiles	18
	Rodents	1173
Congo, Republic Of	* Not Yet Typed	98
	Bats	1803
	Birds	13
	Other Mammals	3398
	Primates	960
	Reptiles	27
	Rodents	1579
Gabon	Other Mammals	100
	Primates	15
	Rodents	1731
India	Bats	394
Lao Peoples Democratic Republic	Bats	3519
	Birds	38
	Other Mammals	642
	Primates	23
	Rodents	5314
	Unknown	8
Malaysia	Bats	1021
	Birds	50
	Other Mammals	2069
	Primates	7199
	Reptiles	28
	Rodents	920
Mexico	Bats	7756
	Other Mammals	24
	Rodents	181
Peru	Bats	283
	Birds	1256
	Other Mammals	272
	Primates	4951
	Reptiles	240
	Rodents	1326
	Unknown	1
Rwanda	* Not Yet Typed	259
	Bats	561
	Other Mammals	40
	Primates	1276

Tanzania, United Republic Of	Rodents	178
	Unknown	70
	* Not Yet Typed	470
	Bats	523
	Other Mammals	1152
Thailand	Primates	75
	Rodents	2262
	Unknown	50
	* Not Yet Typed	213
	Bats	718
Uganda	Other Mammals	1
	Rodents	41
	Bats	681
	Birds	11
	Other Mammals	422
Vietnam	Primates	773
	Rodents	1328
	Birds	161
	Other Mammals	320
	Reptiles	4
Name: SampleDate, Length: 108	Rodents	433
	Unknown	98

## Counts by animal

- Sample types grouped by individual animal



```

Out[291]: <class 'pandas.core.frame.DataFrame'>
Index: 26677 entries, 100000 to Zirikana.2011-01-03
Data columns:
Anal swab                11 non-null values
Biopsy                  2754 non-null values
Blood (whole)           2645 non-null values
Blood Drop              38 non-null values
Blood Drop (Whatman 903) 583 non-null values
Blood Swab              1 non-null values
Blood cells (pellet)    11 non-null values
Blood clot              2014 non-null values
Blood drop (FTA)        3196 non-null values
Blood drop (Nobuto)     29 non-null values
Blood drop (protein saver) 7919 non-null values
Blood smear             877 non-null values
Blood smear (thick)     54 non-null values
Bone                    1 non-null values
Buffy coat              1220 non-null values
Buffy coat + RBC        47 non-null values
Cells (no media)        10 non-null values
Cloacal swab            1285 non-null values
Combined choanal-cloacal swab 301 non-null values
External parasite       1164 non-null values
Feather                 259 non-null values
Feces                   2925 non-null values
Feces (Pool)            1 non-null values
Feces/Rectal swab      1341 non-null values
Fetus                   32 non-null values
Genital swab            624 non-null values
Hair clipping           402 non-null values
Internal parasite       82 non-null values
Lung swab               3 non-null values
Milk                    5 non-null values
Mixed Sample            84 non-null values
Nasal swab              556 non-null values
Ocular swab             1 non-null values
Ophthalmic swab         3 non-null values
Oral swab               12815 non-null values
Oropharyngeal swab     1374 non-null values
Parasite_sample_(specimen) 1 non-null values
Plasma                  1590 non-null values
Rectal swab             9692 non-null values
Red blood cells         3234 non-null values
Red blood cells (clot)  20 non-null values
Saliva                  10 non-null values
Seminal vesicle         3 non-null values
Serum                   6086 non-null values
Skin                    64 non-null values
Smear, thick            135 non-null values
Smear, thin             137 non-null values
Tissue                  4392 non-null values
Tracheal swab           24 non-null values
Urine                   1067 non-null values
Urine/Urogenital swab  1592 non-null values
Whatman paper           403 non-null values
Whisker Follicles       1 non-null values
Whole Animal            2 non-null values
Wing punch              18 non-null values
dtypes: float64(55)

```

- Number of Animals Sampled

```
Out[292]: 26677
```

- Mean number of samples per animal

```
Out[293]: 5.313341080331371
```

- Animals by organization

```

Out[294]: <class 'pandas.core.frame.DataFrame'>
Index: 26839 entries, 100000 to Zirikana.2011-01-03
Data columns:
EHA                7093 non-null values
GVFI               9831 non-null values
Smithsonian        174 non-null values
UC Davis           1726 non-null values
WCS                8015 non-null values
dtypes: float64(5)

```

- Animals by Taxa

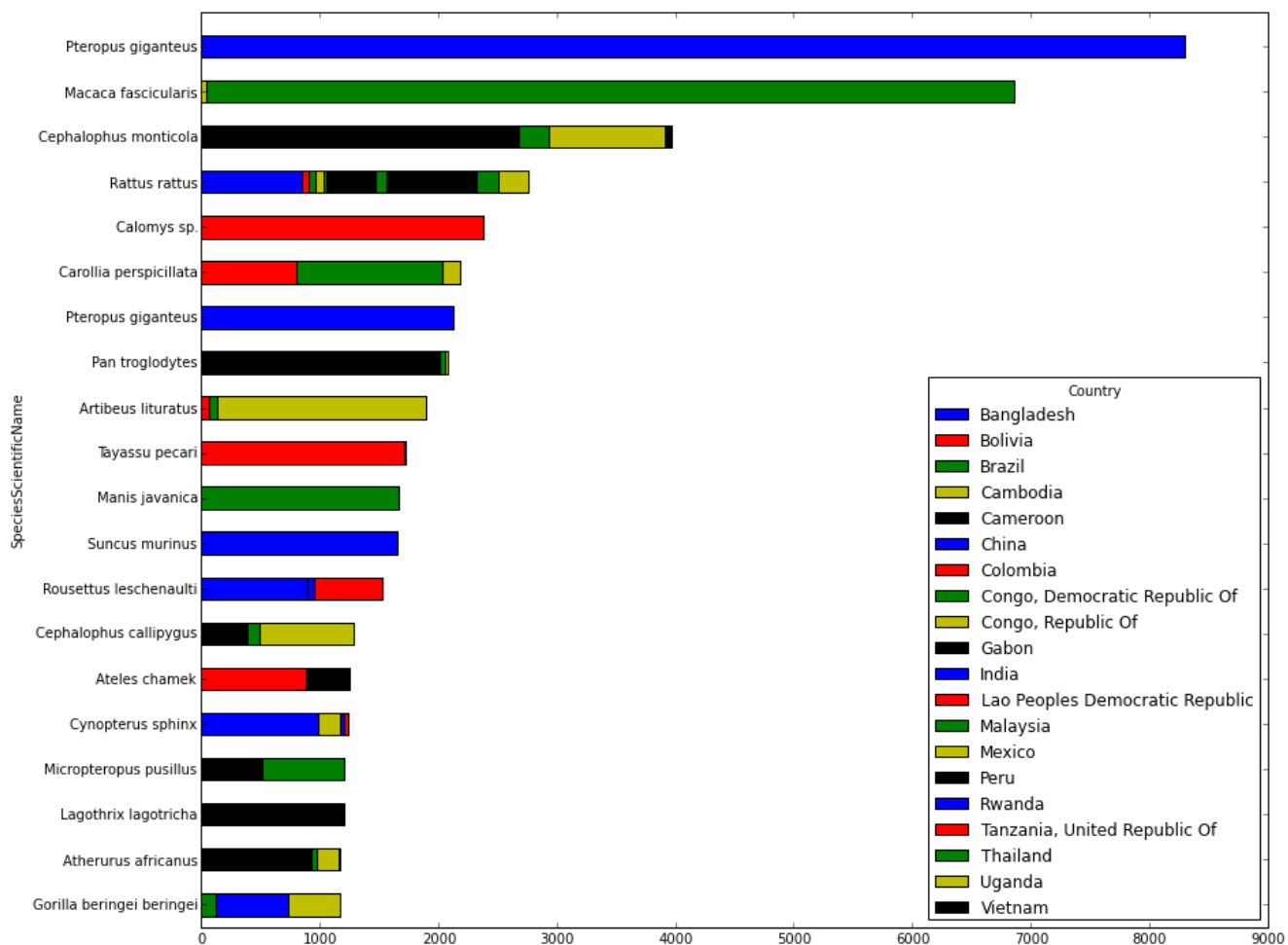
```
Out[295]: <class 'pandas.core.frame.DataFrame'>
Index: 26839 entries, 100000 to Zirikana.2011-01-03
Data columns:
* Not Yet Typed      903  non-null values
Bats                  8137 non-null values
Birds                 1709 non-null values
Other Mammals         4990 non-null values
Primates              4083 non-null values
Reptiles              381  non-null values
Rodents               5618 non-null values
Unknown               1018 non-null values
dtypes: float64(8)
```

- Animals by country

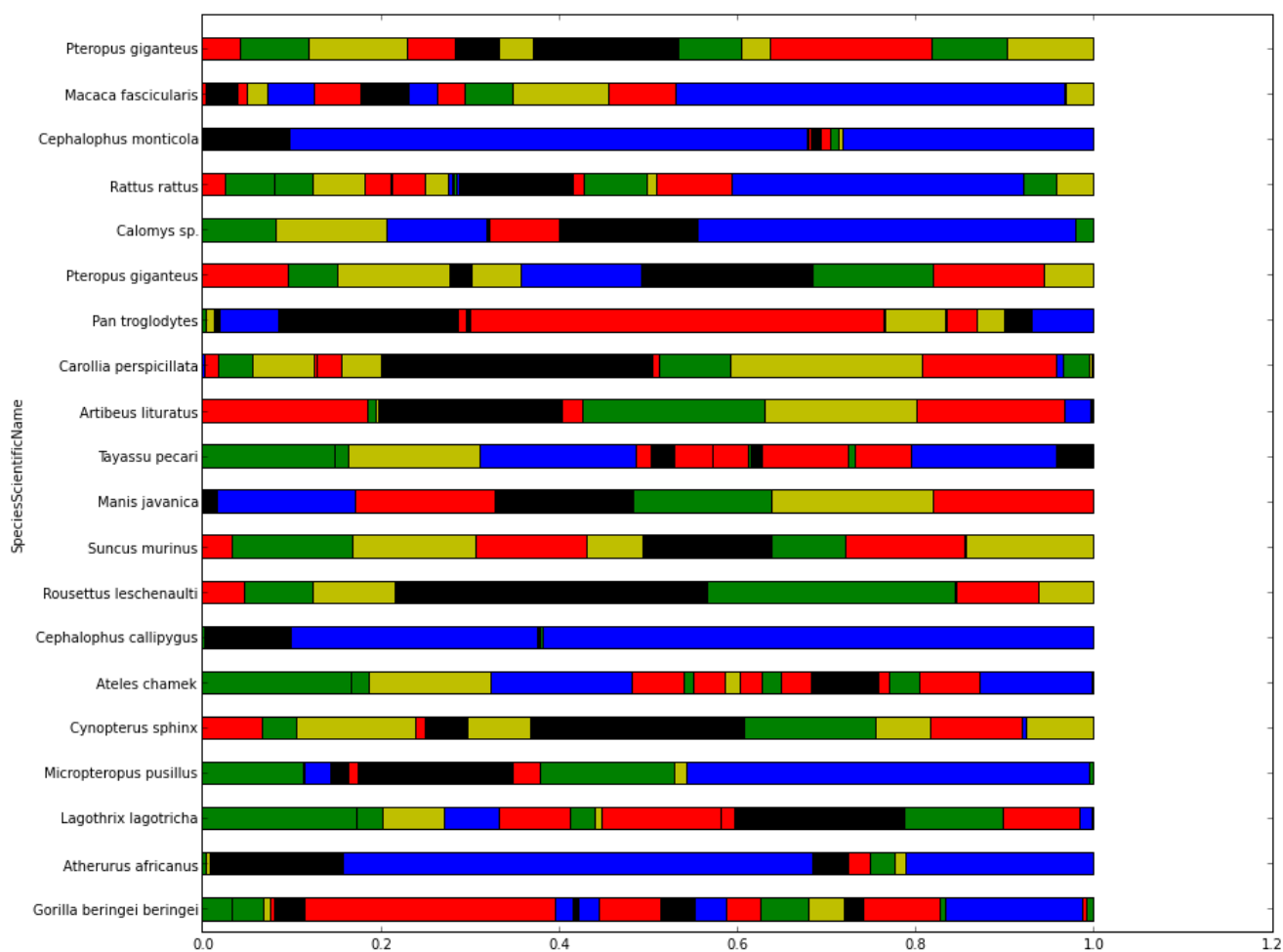
```
Out[296]: <class 'pandas.core.frame.DataFrame'>
Index: 26839 entries, 100000 to Zirikana.2011-01-03
Data columns:
Bangladesh           2018 non-null values
Bolivia              1313 non-null values
Brazil               1022 non-null values
Cambodia             1764 non-null values
Cameroon             6644 non-null values
China                1227 non-null values
Colombia              60  non-null values
Congo, Democratic Republic Of 2339 non-null values
Congo, Republic Of   1007 non-null values
Gabon                870  non-null values
India                59  non-null values
Lao Peoples Democratic Republic 2197 non-null values
Malaysia             737  non-null values
Mexico              1610 non-null values
Peru                 1652 non-null values
Rwanda               327  non-null values
Tanzania, United Republic Of 747  non-null values
Thailand             332  non-null values
Uganda               613  non-null values
Vietnam              301  non-null values
dtypes: float64(20)
```

## Tranches by prevalence

Out[307]: <matplotlib.axes.AxesSubplot at 0x10b7595d0>



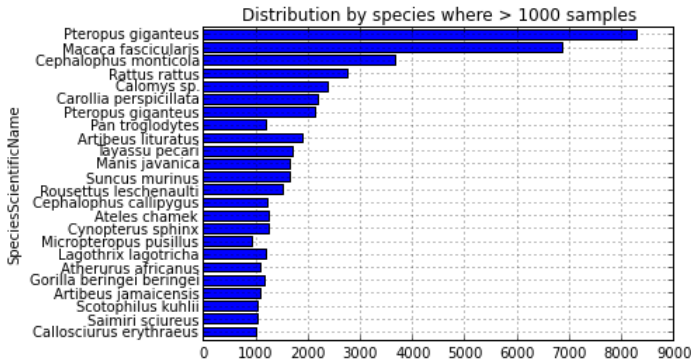
Out[310]: <matplotlib.legend.Legend at 0x10abce890>



```
Out[314]: Index([Malacomys, Megaerops niphanae, Triaenops persicus, Cercocebus agilis, Praomys jacksoni, Alouatta seniculus, Tadari  
brasilienis, Aratinga mitrata, Melopsittacus undulatus, Leopoldamys sabanus, Eonycteris spelaea, Leopoldamys edwardsi, M  
musculoides, Gorilla gorilla gorilla, Cricetomys emini, Macroglossus sobrinus, Nandinia binotata, Mops condylurus, Nasua  
Praomys sp., Artibeus phaeotis, Saguinus fuscicollis, Carolia brevicauda, Syncerus Caffer Nanus, Rhizomys sinensis, Rhi  
sp., Dremomys rufigenis, Bandicota indica, Chaerephon pumila, unknown, Lophuromys, Gorilla beringei graueri, Cebus albif  
Hipposideros gigas, Manis tricuspis, Praomys, Paradoxurus hermaphroditus, Crocidura, Cercopithecus ascanius, Rhinolophus  
Alouatta sara, Cercopithecus mona, Myonycteris torquata, Madoqua kirkii, Rousettus sp., Artibeus planirostris, Mastomys,  
Mandrillus sphinx, Papio hamadryas, Calidris pusilla, Pecari tajacu, Mandrillus leucophaeus, Oryzomys sp., Caiman yacare,  
Porzana fusca, Rousettus leschenaultii, Cebus apella, Cercopithecus cephus, Callosciurus inornatus, Megaloglossus woermar  
musculus, Taphozous melanopogon, Hypsignathus monstrosus, Bandicota bengalensis, Hipposideros caffer, Hystrix brachyura,  
Molossus molossus, Rattus norvegicus, Cephalophus dorsalis, Aratinga acuticaudata, Eidolon helvum, Mastomys natalensis,  
Cynopterus brachyotis, Carolia sowellii, Rousettus leschenaultii , Megaderma lyra, Rattus sp., Cynopterus Sphinx, Sus sci  
Sturnira lilium, Cebus libidinosus, Cercopithecus nictitans, Epomops franqueti, Niviventer fulvescens, Papio anubis, Glos  
soricina, Callosciurus erythraeus, Saimiri sciureus, Scotophilus kuhlii, Artibeus jamaicensis, Gorilla beringei beringei,  
Atherurus africanus, Lagotrix lagotricha, Micropteropus pusillus, Cynopterus sphinx, Ateles chamek, Cephalophus callipyg  
Rousettus leschenaulti, Suncus murinus, Manis javanica, Tayassu pecari, Artibeus lituratus, Pan troglodytes, Pteropus gi  
, Carolia perspicillata, Calomys sp., Rattus rattus, Cephalophus monticola, Macaca fascicularis, Pteropus giganteus],  
dtype=object)
```

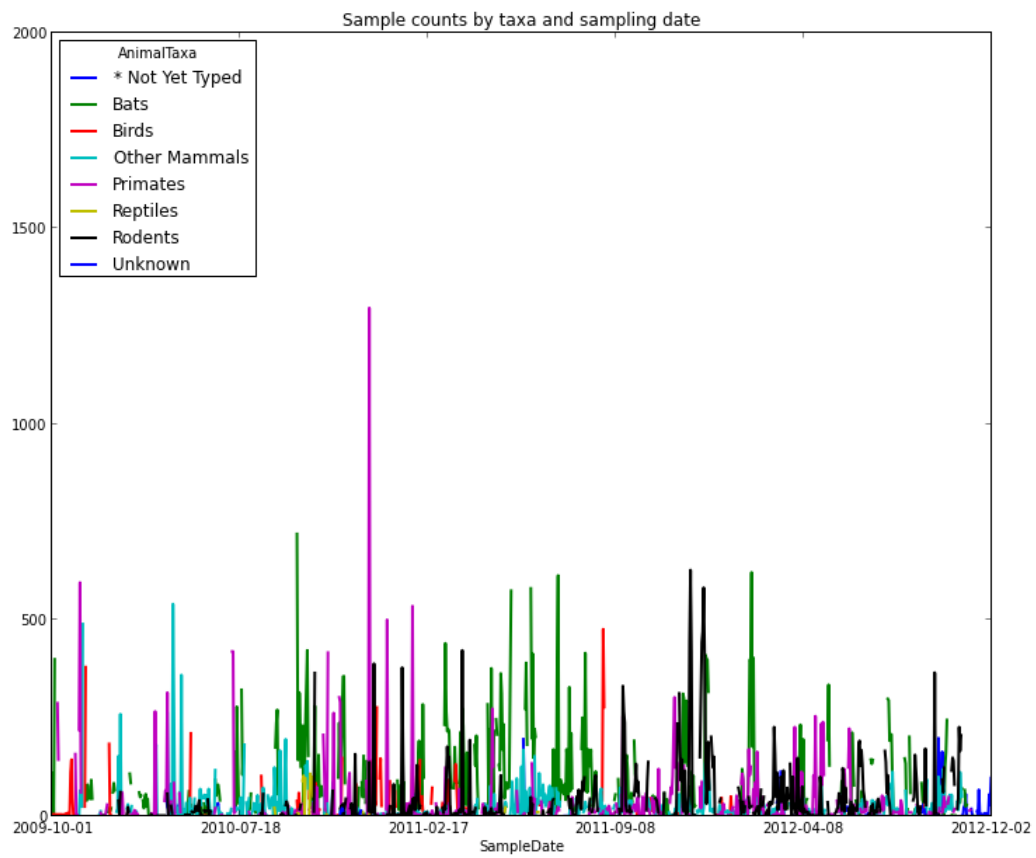
Species with >1000 samples

```
Out[319]: <matplotlib.axes.AxesSubplot at 0x129f8cc50>
```

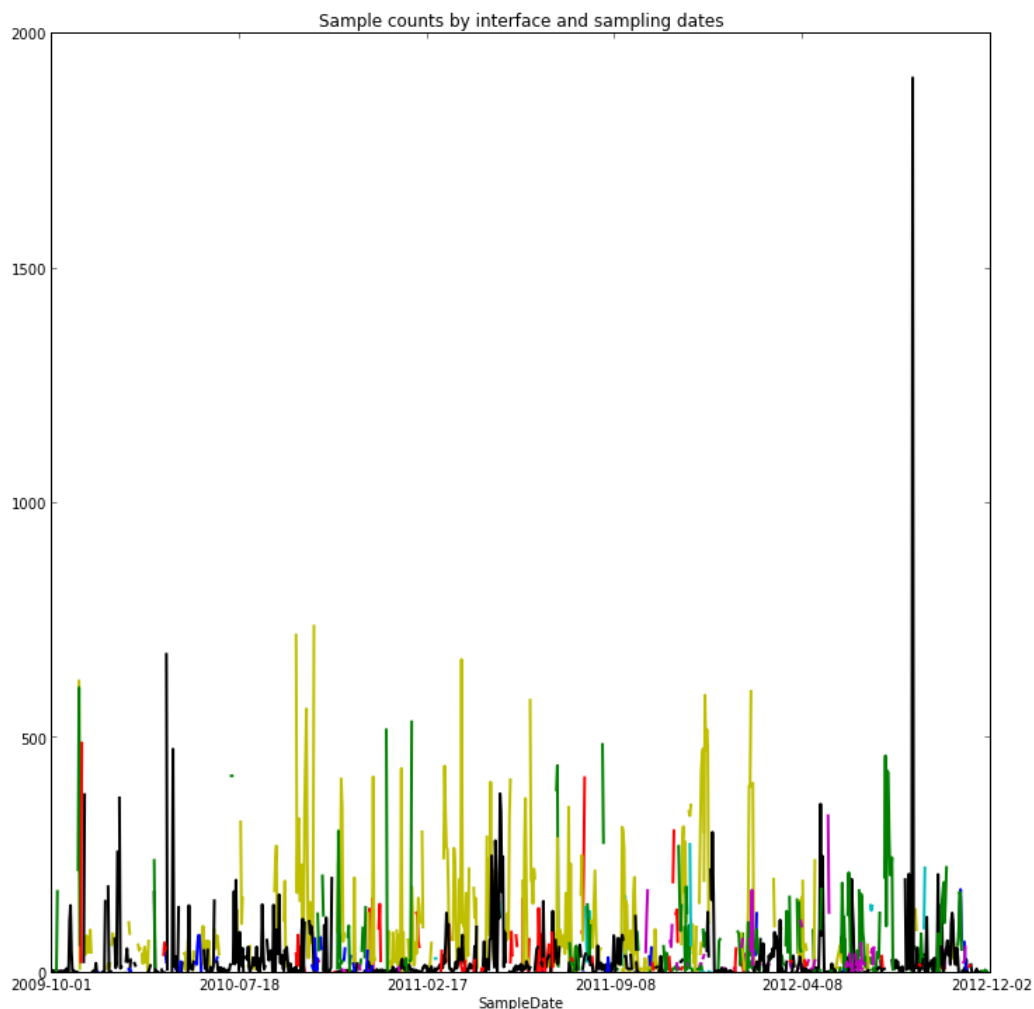
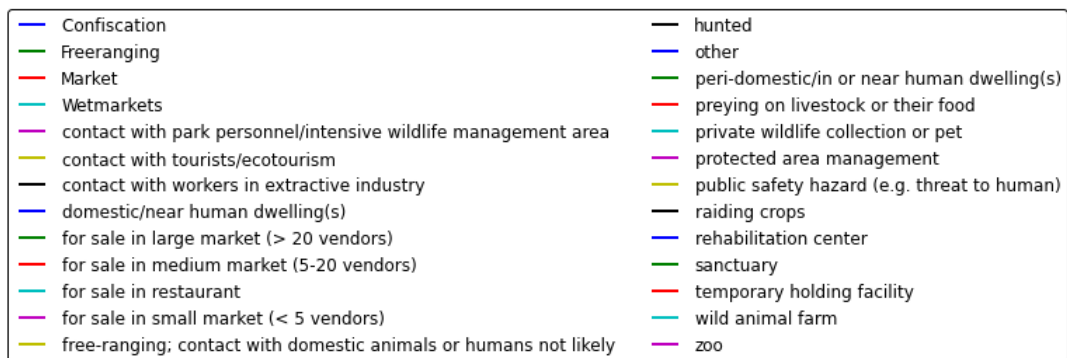


Time Series

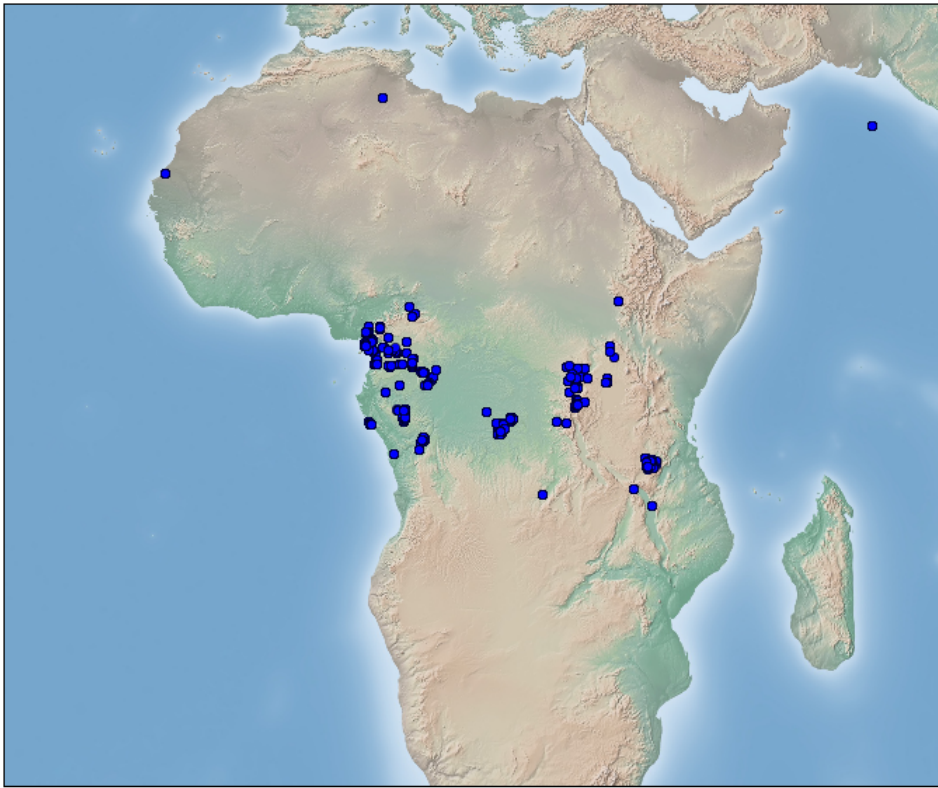
Out[322]: <matplotlib.axes.AxesSubplot at 0x129edd510>



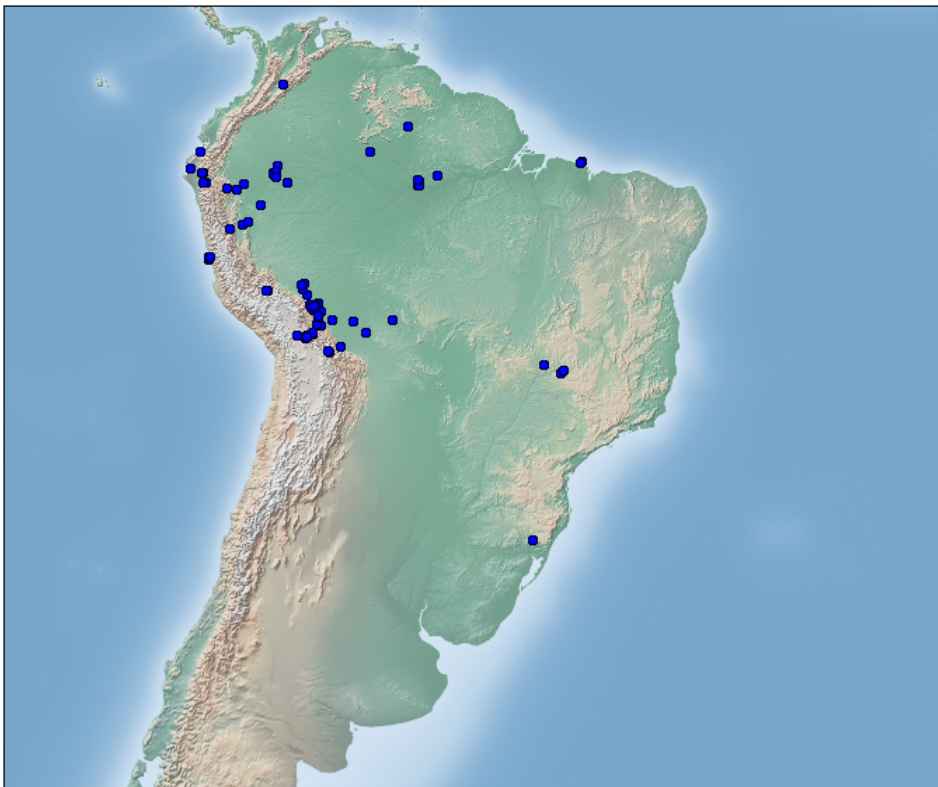
Out[324]: <matplotlib.legend.Legend at 0x10ae74190>



Sampling sites: Africa

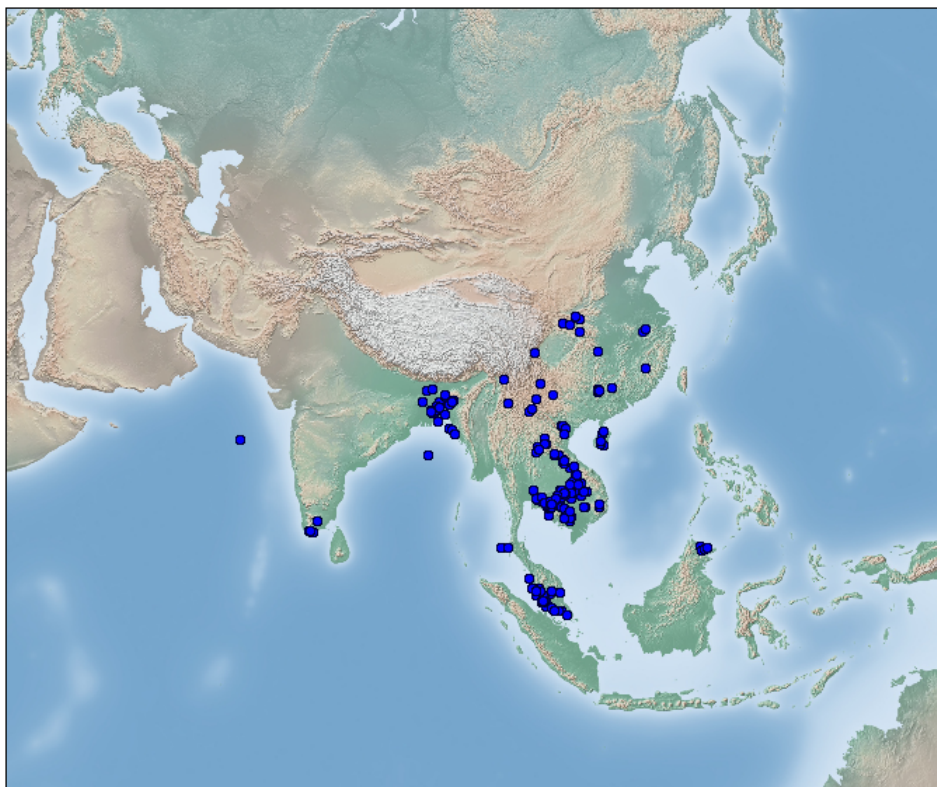


**Sampling Sites: South America**

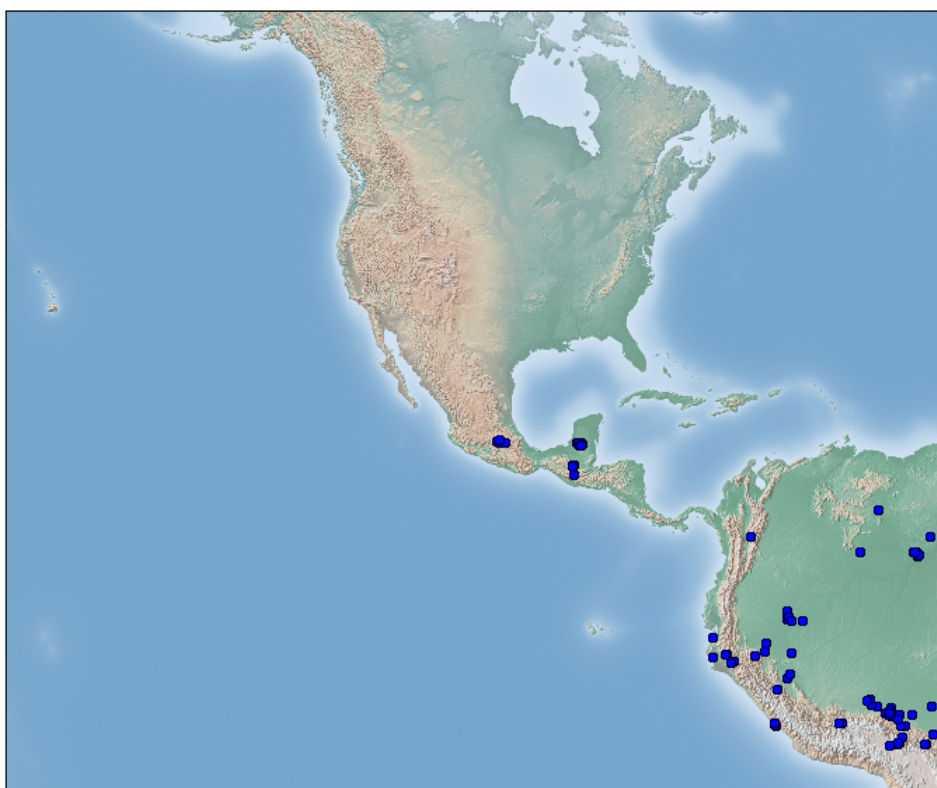


**Sampling Sites: Asia**





**Sampling Sites: Americas**



**Spatial Validation**

Out[330]:

	Latitude	Longitude
count	141505.000000	141505.000000
mean	51.295507	155.126959
std	5063.469931	14149.684793
min	-90.294580	-99.469722
25%	-3.125900	-48.083394
50%	3.625400	18.642920
75%	18.116220	90.040500
max	549872.000000	1536601.000000

- Number of valid longitudes

Out[332]: 141493

- Number of valid latitudes

Out[333]: 140411

- Difference between valid longs and lats

Out[334]: 1082

**. of longitudes out-of-range (<-181 or >181)**

Out[336]: 12

**. of latitudes out-of-range (<-91 or > 91)**

Out[337]: 541

**Distribution of Longitudes (X) and Latitudes (Y)**

