**GBIF (Global Biodiversity Information Facility) Occurrences**

**and Mass Outage Data**

GBIF: What we have so far:

Data for rodents and raptors (Accipitors). We could consider adding data for other orders of raptors and also woodpecker occurrences since these species also have been known to influence outages through degrading wooden utility poles and directly damaging equipment (Polat et al. 2016).

The data all comes from the GBIF database, but GBIF itself sources data from multiple credible occurrence record databases before combining them into a single tabular format. The three main source databases that make up the data we downloaded are eBird (Cornell), iNaturalist, and NatureServe (Natural Heritage Databases). The raptor data has records from all three sources and the rodent data is all from iNaturalist. Depending on which data source the record came from, the field information in GBIF differs. The data sources differ in the ways they format IDs and field codes and they also differ in their data availability. Some records therefore do not have data in certain fields that other records do. How these issues were addressed is documented below.

Datasets:

Rodent\_Occurrences\_cleaned.csv

Rodents (Order: Rodentia)

Temporal Range: 2013-2018

Total Records: 2,147

Temporal Resolution: minutes; 2,038 records with time stamps (~95% of records)

Spatial resolution: lat long coordinates; 2,135 records with Town IDs (~99.5% of data), 525 with Street IDs (~24.5% of data)

Source Data: iNaturalist-Research Grade Observations (2147 records)

|  |  |
| --- | --- |
| **Field Name** | **Definition** |
| gbifID | Unique identifier for the record in the GBIF database |
| Publisher | Organization that provided the data record (source database owner). |
| References | Direction to original observation from the source database (combination of “identifier” and “references” fields) |
| institutionCode | Unique code for the source database organization. |
| datasetName | Source database name |
| basisOfRecord | Method of observation/documentation of the species |
| informationWithheld | Any information withheld about the observation (often for legal protection of a species) |
| OccurrenceRemarks | Observation notes |
| East\_time | Time the species was observed (24 hr Eastern Time, adjusts standard or daily depending on time of year) (Combination of “verbatimEventDateD” and “eventTime” fields) |
| UTC\_Time | Time the species was observed (24 hr coordinated universal time) (Combination of “verbatimEventDate” and “eventTime” fields) |
| Date | Date the observation was made (YYY-MM\_DD) |
| Year | Year the observation was made |
| Month | Month the observation was made |
| Day | Day the observation was made |
| countryCode | Country the observation was made |
| County | County the observation was made |
| Locality | Fine details on observation location. Varies from townships, street names, to property names. (Combination of “locality” and “verbatimLocality” fields) |
| decimalLatitude | Latitude |
| decimalLongitude | Longitude |
| coordinateUncertaintyInMeters | Coordinate uncertainty in meters |
| Order | Taxonomic order |
| Family | Taxomonic family |
| Genus | Taxonomic genus |
| Subgenus | Taxonomic sub-genus |
| specificEpithet | Taxonomic species |
| infraspecificEpithet | Taxonomic sub-species |
| taxonRank | The taxonomic level that the species was identified to (order, family, genus, species, subspecies) |
| Elevation | Elevation above sea-level that the observation was made |
| Issue | Notes on any assumptions or adjustments made to the occurrence when incorporating it into the database. (Mostly spatial adjustments). |
| hasCoordinate | 1- has lat long coordinate data; 0- does not have lat long coordinate data |
| hasGeospatialIssues | 1 -Geospatial issues associated with the record may increase uncertainty; 0- no geospatial issues. |
| orderKey | Unique numerical ID associated with the species’ taxonomic order |
| familyKey | Unique numerical ID associated with the species’ taxonomic family |
| genusKey | Unique numerical ID associated with the species’ taxonomic genus |
| speciesKey | Unique numerical ID associated with the species’ taxonomic species |
| species | Scientific name |
| duplicate | 1. Indicates record is a duplicate observation, 0- indicates record is a unique observation |
| hasTime | 1. Indicates record has temporal resolution to minutes, 0- record does not have a temporal resolution past date |
| individualCount | abundance/number of individuals observed; only consistently available with eBird associated records |

Adjustments made during data cleaning:

* Removed all empty fields. Included:
  + abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo contributor, coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted, description, educationLevel, extent, format, hasFormat, hasPart, hasVersion, mediator, medium, modified, provenance relation, replaces, requires, source, spatial, subject, tableOfContents, temporal, title, type, valid, institutionID, collectionID, datasetID, ownerInstititionCode, dataGeneralizations, dynamicProperties, organismQuantity, organismQuantityType, sex, lifeStage, reproductiveCondition, behavior, establishmentMeans, occurrenceStatus, preparations, disposition, associatedReferences, associatedSequences, associatedTaxa, otherCatalogNumbers, occurrenceRemarks, organismID, organismName, organismScope, associatedOccurrences, associatedOrganisms, previousIdentifications, organismRemarks, materialSampleID, eventide, parentEventID, recordnumber, startDayOfYear, endDayOfYear, habitat, samplingProtocol, samplingEffort, sampleSizeValue, sampleSizeUnit, fieldNotes, locationID, higherGeographyID, higherGeography, waterbody, islandGroup, island, municipality, verbatimElevation, verbatimDepth, minimumDistanceAboveSurfaceInMeters, maximumDistanceAboveSurfaceInMeters, locationAccordingTo, locationRemarks, coordinatePrecision, pointRadiusSpatialFit, verbatimCoordinateSystem, verbatimSRS, footprintWKT, footprintSRS, footprintSpatialFit, georeferencedBy, georeferencedDate, georeferenceProtocol, georeferenceSources, georeferenceVerificationStatus, georeferenceRemarks, geologicalContextID, earliestEonOrLowestEonothem, latestEonOrHighestEonothem, earliestEraOrLowestErathem, latestEraOrHighestErathem, earliestPeriodOrLowestSystem, latestPeriodOrHighestSystem, earliestEpochOrLowestSeries, latestEpochOrHighestSeries, earliestAgeOrLowestStage, latestAgeOrHighestStage, lowestBiostratigraphicZone, highestBiostratigraphicZone, lithostratigraphicTerms, group, formation, member, bed, identificationID, identificationQualifier typeStatus, identifiedBy, dateIdentified, identificationReferences, identificationVerificationStatus, identificationRemarks, parentNameUsageID, originalNameUsageID, nameAccordingToID, namePublishedInID, taxonConceptID, acceptedNameUsage, parentNameUsage, originalNameUsage, nameAccordingTo, namePublishedIn, namePublishedInYear, higherClassification, verbatimTaxonRank, vernacularName, nomenclaturalCode, nomenclaturalStatus, taxonRemarks, elevationAccuracy, depth, depthAccuracy, distanceAboveSurface, distanceAboveSurfaceAccuracy, repatriated, relativeOrganismQuantity.
* Fields removed due to inconsistent records, a lack of applicability, or redundancy:
  + license, rights (inconsistent), rightsholder (inconsistent), occurrenceID (record IDs differ between data sources), catalogNumber (Unique identifiers that differ between data sources), fieldNUmber (Natureserve IDs, not applicable to other data sources), eventRemarks (Only for Natureserve data), Continent, identificationID (Unique identifier only for iNaturalist data), dateIdentified (only iNaturalist data), identificationRemarks (iNaturalist data only), taxonID (only iNaturalist data), acceptedNameUsageID (redundant of speciesKey field), scientificName (redundant of species field), scientificNameID (redundant of speciesKey field), Kingdom, Phylum, Class, DatasetKey (UUID key for dataset registered in GBIF), publishingCountry (redundant of countryCode field), lastInterpreted (time stamp of last time the record has been reinterpreted by GBIF), mediatype (format of record), taxonKey (redundant of speciesKey field), acceptedTaxonKey (redundant of speciesKey field), kingdomKey, PhylumKey, classKey, acceptedScientificName (scientific name with source reference, redundant of species field), typifiedName, Protocol, lastParsed
* Fields edited and combined:
  + “References”- combined data from fields “identifier” and “references”. This brought together all catalog numbers to link each record to the original record in its source database.
  + “Time\_EST”- referenced the verbatimEventDate field for EST time stamps (only available for iNaturalist data). Any missing time stamps were converted from the UTC time field (eventTime). Any EDT times were also converted to EST. All times were then converted to 24 hr time (hh:mm:ss).
  + “Time\_UTC”- referenced the eventTime field for UTC time stamps (only available for iNautralist data). Any missing time stamps were converted from the EST time field (verbatimEventDate). All times were then converted to 24 hr time (hh:mm:ss).
  + Date- combined the year, month, and day fields separated by “-“
* Other considerations to note:
  + Times are also available in the original eBird data, but they were lost when the records were brought into GBIF. We are working on obtaining this data and corresponding the observation times to the catalog number in the “references” field. Observation times are not available for NatureServe data, but these records make up a smaller percentage of the GBIF database.
  + Lat/Long coordinates are not available for NatureServe data. The smallest spatial resolution for these records is down to state counties.
  + Empty field values were given “NA” values.

Duplicate Records Removed

Rodent data:

The following gbifIDs were duplicate records listed as separate observations on iNaturalist-

2465138186, 1948738627, 1933549627, 1933548876, 1932358464, 1932358038, 1932357861, 1890067168, 1847617165, 1836697139, 1831191705, 1831096906, 1562892939, 1143562001

Record Filtering Columns Added

Two column fields already existed “hasCoordinate” and “hasGeospatialIssues”. These columns were “TRUE” “FALSE” binary variables and TRUE values were changed to 1, FALSE values to 0. To filter out records without geospatial coordinates or with geospatial issues, select the subset of records with hasCoordinate =1 and hasGeospatialIssues = 0. These two fields will apply to the raptor data. The rodent data has complete coordinate coverage.

Two columns were added “duplicate” and “hasTime”. These follow the same rules where 1 = TRUE and 0 = FALSE. In this case records with 1 in the duplicate field are duplicates. To filter out duplicate records, remove the subset of records with duplicate = 1 or select the subset of records with duplicate = 0 records. To only select records with minute resolution time stamps, select the subset of records with hasTime = 1.

Exploratory Summaries of Rodent Data

Total species: 18

Highlighted species are known to cause issues with electrical equipment above and underground. \* make up greater than 90% of records.

|  |  |
| --- | --- |
| **Scientific Name** | **Common Name** |
| *Castor canadensis\** | American beaver |
| *Erethizon dorsatus* | North American porcupine |
| *Glaucomys sabrinus* | Northern flying squirrel |
| *Glaucomys volans* | Southern flying squirrel |
| *Marmota monax\** | Groundhog |
| *Microtus pennsylvanicus* | Meadow vole |
| *Microtus pinetorum* | Woodland vole |
| *Mus musculus* | House mouse |
| *Myodes gapperi* | Southern red-backed vole |
| *Napaeozapus insignis* | Woodland jumping mouse |
| *Ondatra zibethicus* | Muskrat |
| *Peromyscus leucopus* | White-footed mouse |
| *Peromyscus maniculatus* | Deer mouse |
| *Rattus norvegicus* | Brown rat |
| *Sciurus carolinensis \** | Eastern gray squirrel |
| *Tamias striatus\** | Eastern chipmunk |
| *Tamiasciurus hudsonicus \** | American red squirrel |
| *Zapus hudsonius* | Meadow jumping mouse |

Datasets:

Rodent\_Occurrences\_townID.csv

**Summarizing occurrences at spatial units:**

An accepted measure of species occurrence data is frequency or density of occurrences. In order to create a measurable and comparable metric out of rodent occurrences, point locations of occurrences were joined to a Massachusetts towns shapefile layer provided with the Massachusetts outage data. Towns will be used as spatial units for further analysis.

Unique town identification numbers (geo\_ids, renamed to TOWN\_ID) were joined to rodent occurrences via a spatial join intersection in QGIS 3. Records that were outside state boundaries were discarded.

Because occurrence points fell outside the Massachusetts state boundary, the original total number of occurrence records differ from the number of records with town IDs.

Number of occurrence records with town correspondence (out of total occurrence records available):

2,148 records total; 2,142 with Town IDs (~99%)

Field names same as above but with additional fields:

|  |  |
| --- | --- |
| **Field Name** | **Definition** |
| TOWN | Adjusted City/town name in the Massachusetts Outage dataset (“actual\_city\_town”) |
| TOWN\_ID | Unique identifier corresponding to each city/town name (renamed from “geo\_id”) |
| POP2010 | US Census city/town human population in 2010 |
| POPWeight | Relative human population size used as a weight for calculating species density/frequency. Accounts for detection bias. (minimum POP2010)/(POP2010 of town x) |
| Area\_km2 | Area of the town/city in square kilometers. |

Datasets:

Raptor\_Occurrences\_cleaned.csv

Raptors (Order: Accipitriformes)

Temporal Range: 2013-2018

Total records: 330,627

Temporal Resolution: minutes; 893 records with time stamps (~0.3% of records)

Spatial resolution: lat long coordinates; 244,625 records with Town IDs (~74%), 117,466 records with Street IDs (~35.5%)

Source Data: iNaturalist-Research Grade Observations (948 records), eBird (329,649 records), NatureServe (30 records)

|  |  |
| --- | --- |
| **Field Name** | **Definition** |
| gbifID | Unique identifier for the record in the GBIF database |
| Publisher | Organization that provided the data record (source database owner). |
| References | Direction to original observation from the source database (combination of “identifier” and “references” fields) |
| institutionCode | Unique code for the source database organization. |
| datasetName | Source database name |
| basisOfRecord | Method of observation/documentation of the species |
| informationWithheld | Any information withheld about the observation (often for legal protection of a species) |
| OccurrenceRemarks | Observation notes |
| East\_time | Time the species was observed (24 hr Eastern Time, adjusts standard or daily depending on time of year) (Combination of “verbatimEventDateD” and “eventTime” fields) |
| UTC\_Time | Time the species was observed (24 hr coordinated universal time) (Combination of “verbatimEventDate” and “eventTime” fields) |
| Date | Date the observation was made (YYY-MM\_DD) |
| Year | Year the observation was made |
| Month | Month the observation was made |
| Day | Day the observation was made |
| countryCode | Country the observation was made |
| County | County the observation was made |
| Locality | Fine details on observation location. Varies from townships, street names, to property names. (Combination of “locality” and “verbatimLocality” fields) |
| decimalLatitude | Latitude |
| decimalLongitude | Longitude |
| coordinateUncertaintyInMeters | Coordinate uncertainty in meters |
| Order | Taxonomic order |
| Family | Taxomonic family |
| Genus | Taxonomic genus |
| Subgenus | Taxonomic sub-genus |
| specificEpithet | Taxonomic species |
| infraspecificEpithet | Taxonomic sub-species |
| taxonRank | The taxonomic level that the species was identified to (order, family, genus, species, subspecies) |
| Elevation | Elevation above sea-level that the observation was made |
| Issue | Notes on any assumptions or adjustments made to the occurrence when incorporating it into the database. (Mostly spatial adjustments). |
| hasCoordinate | 1- has lat long coordinate data; 0- does not have lat long coordinate data |
| hasGeospatialIssues | 1 -Geospatial issues associated with the record may increase uncertainty; 0- no geospatial issues. |
| orderKey | Unique numerical ID associated with the species’ taxonomic order |
| familyKey | Unique numerical ID associated with the species’ taxonomic family |
| genusKey | Unique numerical ID associated with the species’ taxonomic genus |
| speciesKey | Unique numerical ID associated with the species’ taxonomic species |
| species | Scientific name |
| duplicate | 1. Indicates record is a duplicate observation, 0- indicates record is a unique observation |
| hasTIme | 1. Indicates record has temporal resolution to minutes, 0- record does not have a temporal resolution past date |
| individualCount | abundance/number of individuals observed; only consistently available with eBird associated records |

Adjustments made during data cleaning:

* Removed all empty fields. Included:
  + abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo contributor, coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted, description, educationLevel, extent, format, hasFormat, hasPart, hasVersion, mediator, medium, modified, provenance relation, replaces, requires, source, spatial, subject, tableOfContents, temporal, title, type, valid, institutionID, collectionID, datasetID, ownerInstititionCode, dataGeneralizations, dynamicProperties, organismQuantity, organismQuantityType, sex, lifeStage, reproductiveCondition, behavior, establishmentMeans, occurrenceStatus, preparations, disposition, associatedReferences, associatedSequences, associatedTaxa, otherCatalogNumbers, occurrenceRemarks, organismID, organismName, organismScope, associatedOccurrences, associatedOrganisms, previousIdentifications, organismRemarks, materialSampleID, eventide, parentEventID, recordnumber, startDayOfYear, endDayOfYear, habitat, samplingProtocol, samplingEffort, sampleSizeValue, sampleSizeUnit, fieldNotes, locationID, higherGeographyID, higherGeography, waterbody, islandGroup, island, municipality, verbatimElevation, verbatimDepth, minimumDistanceAboveSurfaceInMeters, maximumDistanceAboveSurfaceInMeters, locationAccordingTo, locationRemarks, coordinatePrecision, pointRadiusSpatialFit, verbatimCoordinateSystem, verbatimSRS, footprintWKT, footprintSRS, footprintSpatialFit, georeferencedBy, georeferencedDate, georeferenceProtocol, georeferenceSources, georeferenceVerificationStatus, georeferenceRemarks, geologicalContextID, earliestEonOrLowestEonothem, latestEonOrHighestEonothem, earliestEraOrLowestErathem, latestEraOrHighestErathem, earliestPeriodOrLowestSystem, latestPeriodOrHighestSystem, earliestEpochOrLowestSeries, latestEpochOrHighestSeries, earliestAgeOrLowestStage, latestAgeOrHighestStage, lowestBiostratigraphicZone, highestBiostratigraphicZone, lithostratigraphicTerms, group, formation, member, bed, identificationID, identificationQualifier typeStatus, identifiedBy, dateIdentified, identificationReferences, identificationVerificationStatus, identificationRemarks, parentNameUsageID, originalNameUsageID, nameAccordingToID, namePublishedInID, taxonConceptID, acceptedNameUsage, parentNameUsage, originalNameUsage, nameAccordingTo, namePublishedIn, namePublishedInYear, higherClassification, verbatimTaxonRank, vernacularName, nomenclaturalCode, nomenclaturalStatus, taxonRemarks, elevationAccuracy, depth, depthAccuracy, distanceAboveSurface, distanceAboveSurfaceAccuracy, repatriated, relativeOrganismQuantity.
* Fields removed due to inconsistent records, a lack of applicability, or redundancy:
  + license, rights (inconsistent), rightsholder (inconsistent), occurrenceID (record IDs differ between data sources), catalogNumber (Unique identifiers that differ between data sources), fieldNUmber (Natureserve IDs, not applicable to other data sources), eventRemarks (Only for Natureserve data), Continent, identificationID (Unique identifier only for iNaturalist data), dateIdentified (only iNaturalist data), identificationRemarks (iNaturalist data only), taxonID (only iNaturalist data), acceptedNameUsageID (redundant of speciesKey field), scientificName (redundant of species field), scientificNameID (redundant of speciesKey field), Kingdom, Phylum, Class, DatasetKey (UUID key for dataset registered in GBIF), publishingCountry (redundant of countryCode field), lastInterpreted (time stamp of last time the record has been reinterpreted by GBIF), mediatype (format of record), taxonKey (redundant of speciesKey field), acceptedTaxonKey (redundant of speciesKey field), kingdomKey, PhylumKey, classKey, acceptedScientificName (scientific name with source reference, redundant of species field), typifiedName, Protocol, lastParsed
* Fields edited and combined:
  + “References”- combined data from fields “identifier” and “references”. This brought together all catalog numbers to link each record to the original record in its source database.
  + “East\_Time”- referenced the verbatimEventDate field for EST time stamps (only available for iNaturalist data). Any missing time stamps were converted from the UTC time field (eventTime). All times were then converted to 24 hr time (hh:mm:ss).
  + “UTC\_Time”- referenced the eventTime field for UTC time stamps (only available for iNautralist data). Any missing time stamps were converted from the East time field (verbatimEventDate). All times were then converted to 24 hr time (hh:mm:ss).
  + Date- combined the year, month, and day fields separated by “-“
* Other considerations to note:
  + Times are also available in the original eBird data, but they were lost when the records were brought into GBIF. We are working on obtaining this data and corresponding the observation times to the catalog number in the “references” field. This would increase temporal resolution for raptor data significantly since ebird records make up 99% of the records available. Observation times are not available for NatureServe data, but these records make up a smaller percentage of the GBIF database.
  + Lat/Long coordinates are not available for NatureServe data. The smallest spatial resolution for these records is down to state counties.
  + Empty field values were given “NA” values.

Duplicate Records Removed

Raptor data:

Duplicates were filtered using species name, date, year, day, month, county, decimal longitude and latitude. Until we are able to obtain time resolution data for ebird, these filters are only based on records having the same date (down to the specific day), species, and location data. Only records with temporal resolutions down to days were compared as duplicates. This way records missing dates or only with year values would not be removed. The following gbifIDs were marked as duplicate records listed as separate observations on:

* iNaturalist (n=11)
* eBird(n=93,003)

Record Filtering Columns Added

Two column fields already existed “hasCoordinate” and “hasGeospatialIssues”. These columns were “TRUE” “FALSE” binary variables and TRUE values were changed to 1, FALSE values to 0. To filter out records without geospatial coordinates or with geospatial issues, select the subset of records with hasCoordinate =1 and hasGeospatialIssues = 0. These two fields will apply to the raptor data. The rodent data has complete coordinate coverage.

Two columns were added “duplicate” and “hasTime”. These follow the same rules where 1 = TRUE and 0 = FALSE. In this case records with 1 in the duplicate field are duplicates. To filter out duplicate records, remove the subset of records with duplicate = 1 or select the subset of records with duplicate = 0 records. To only select records with minute resolution time stamps, select the subset of records with hasTime = 1.

Exploratory Summaries of Raptor Data

Total species: 17

Highlighted species are known to cause electrical outages. Species with \* make up 98% of occurrence records out of total annual records within each year (2013-2018).

|  |  |
| --- | --- |
| **Scientific Name** | **Common Name** |
| *Buteo jamaicensis \** | Red-tailed hawk |
| *Pandion haliaetus\** | Osprey |
| *Haliaeetus leucocephalus\** | Bald eagle |
| *Circus hudsonius\** | Northern harrier |
| *Cathartes aura\** | Turkey vulture |
| *Accipiter striatus\** | Sharp-shinned hawk |
| *Buteo lagopus* | Rough-legged buzzard |
| *Accipiter cooperii\** | Cooper’s hawk |
| *Ictinia mississippiensis* | Mississippi kite |
| *Coragyps atratus* | Black vulture |
| *Buteo lineatus\** | Red-shouldered hawk |
| *Accipiter gentilis* | Northern goshawk |
| *Buteo platypterus\** | Broad-winged hawk |
| *Aquila chrysaetos* | Golden eagle |
| *Elanoides forficatus* | Swallow-tailed kite |
| *Buteo albonotatus* | Zone-tailed hawk |
| *Buteo swainsoni* | Swainson’s hawk |

Datasets:

Raptor\_Occurrences\_townstreetID.csv

**Summarizing occurrences at spatial units:**

An accepted measure of species occurrence data is frequency or density of occurrences. In order to create a measurable and comparable metric out of raptor occurrences, point locations of occurrences were joined to a Massachusetts towns shapefile layer provided with the Massachusetts outage data. Towns will be used as spatial units for further analysis.

Unique town identification numbers (geo\_ids, renamed to TOWN\_ID) were joined to raptor occurrences via a spatial join intersection in QGIS 3. Records that were outside state boundaries were discarded.

Because occurrence points fell outside the Massachusetts state boundary, the original total number of occurrence records differ from the number of records with town IDs. In addition, this process removed records that did not have spatial resolution to latitude/longitude or city (NatureServe records).

Number of occurrence records with town and/or street correspondence (out of total occurrence records available):

330,628 records total; 329,505 with Town IDs (~99.6%)

Field names same as above but with additional fields:

|  |  |
| --- | --- |
| **Field Name** | **Definition** |
| TOWN | Adjusted City/town name in the Massachusetts Outage dataset (“actual\_city\_town”) |
| TOWN\_ID | Unique identifier corresponding to each city/town name (renamed from “geo\_id”) |
| POP2010 | US Census city/town human population in 2010 |
| POPWeight | Relative human population size used as a weight for calculating species density/frequency. Accounts for detection bias. (minimum POP2010)/(POP2010 of town x) |
| Area\_km2 | Area of the town/city in square kilometers. |
| Observer\_ID | Extrapolated from the “recordedby” field from GBIF. Unique identifier for the observer who made the occurrence observation. Observer IDs were used to generate observer weights for unique town IDs by counting the number of unique observers per town and calculating a similar weight as POPweight (minimum number of observers)/(number of observers in town x). |

Datasets:

MA\_outages\_townID.csv

The Massachusetts outage data original town names were inconsistent with actual state cities and towns due to some discrepancies on state borders with outages within cities in neighboring states. Some names were neighborhood names and needed to be adjusted to the city that contained them.

With both the outage records and species records joined to the same spatial units (town IDs) we can begin summarizing frequency counts and densities at each spatial unit over time.

Crosswalk for Town names. Outage data included City.Town names that classified as neighborhoods. Those were translated to the City each neighborhood is a part of. Some neighborhood names no longer existed in city records and some city names had abbreviations.

|  |  |
| --- | --- |
| **Original Town Name** | **Adjusted Town Name** |
| BOSTON DOWNTOWN | BOSTON |
| BOURNE PLYMOUTH | BOURNE |
| BRIGHTON | BOSTON |
| CHARLESTOWN | BOSTON |
| DORCHESTER | BOSTON |
| E. BRIDGEWATER | EAST BRIDGEWATER |
| EAST BOSTON | BOSTON |
| ENFIELD | BOSTON |
| EPSOM | HAMPDEN |
| HYANNIS | BARNSTABLE |
| HYDE PARK | BOSTON |
| MT. WASHINGTON | MOUNT WASHINGTON |
| NEW MARLBORO | NEW MARLBOROUGH |
| NORTH TISBURY | WEST TISBURY |
| ROXBURY | BOSTON |
| SANDWICH PLYMOUTH | SANDWICH |
| SOUTH BOSTON | BOSTON |
| WEST ROXBURY | BOSTON |

Towns not in the outage data. Towns were added to outage data with empty (NA) records in order to represent a complete spatial distribution.

|  |
| --- |
| **Town Names added to outages** |
| BELMONT |
| BERKLEY |
| BOYLSTON |
| BRAINTREE |
| DANVERS |
| HOLDEN |
| HUDSON |
| HULL |
| IPSWICH |
| LYNNFIELD |
| MANSFIELD |
| MARBLEHEAD |
| MERRIMAC |
| NORTH ATTLEBOROUGH |
| NORTH READING |
| OAKHAM |
| PAXTON |
| PEABODY |
| PRINCETON |
| RAYNHAM |
| ROWLEY |
| SHREWSBURY |
| SOUTH HADLEY |
| TAUNTON |
| TEMPLETON |
| WAKEFIELD |
| WELLESLEY |
| WEST BOYLSTON |
| GROTON |
| GOSNOLD |
| GEORGETOWN |
| NEW BRAINTREE |
| MIDDLETON |
| BOXBOROUGH |
| LITTLETON |

Initial Exploration of Data Composition:

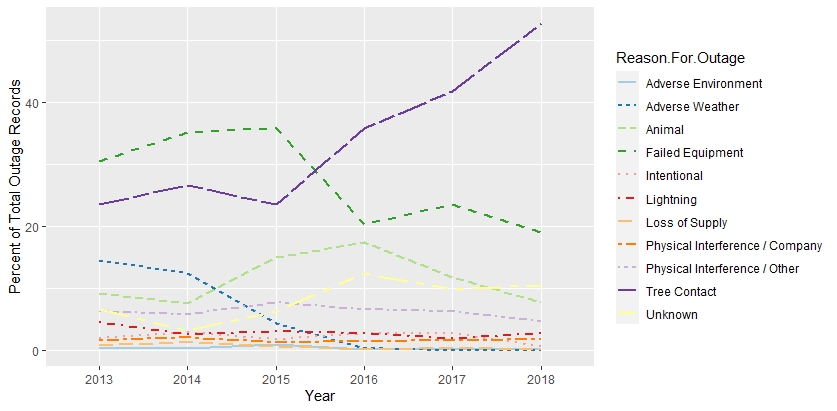


Figure 1. The composition of outage causes across years (2013-2018) with each outage cause as a percent of the total number of outages recorded each year. A majority of causes have declined over the time period, but Tree Contact has increased. Animal and Unknown causes gradually increased over the time period, cresting in 2016. Tree Contact, Failed Equipment, Unknown, and Animal are within the top four causes of outages from 2013-2018 and account for just under 90% of outages in 2018. Adverse weather and Physical Interference/Other were larger contributors from 2013-2015, but have declined since then.

Table 1. The composition of outage causes in 2018. Total outages in 2018 for each outage cause as a percentage of the total number of outages recorded in 2018.

|  |  |  |
| --- | --- | --- |
| Outage Cause | Number of Outages | Percent of Total Outages in 2018 |
| Tree Contact | 22645 | 52.7 |
| Failed Equipment | 8170 | 19.04 |
| Unknown | 4498 | 10.4 |
| Animal | 3279 | 7.6 |
| Physical Interference / Other | 2002 | 4.6 |
| Lightning | 1128 | 2.6 |
| Physical Interference / Company | 743 | 1.7 |
| Intentional | 292 | 0.68 |
| Adverse Environment | 67 | 0.15 |
| Loss of Supply | 63 | 0.14 |
| Adverse Weather | 8 | 0.01 |

Figure 2. Total number of outage records per year from 2013-2018 in red and total number of Animal caused outages per year in blue. Both increase with time, the total number of records being largely driven by tree contact outages.

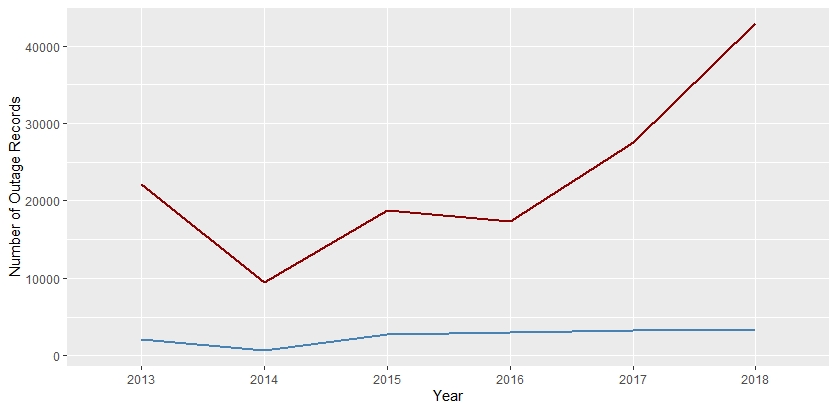
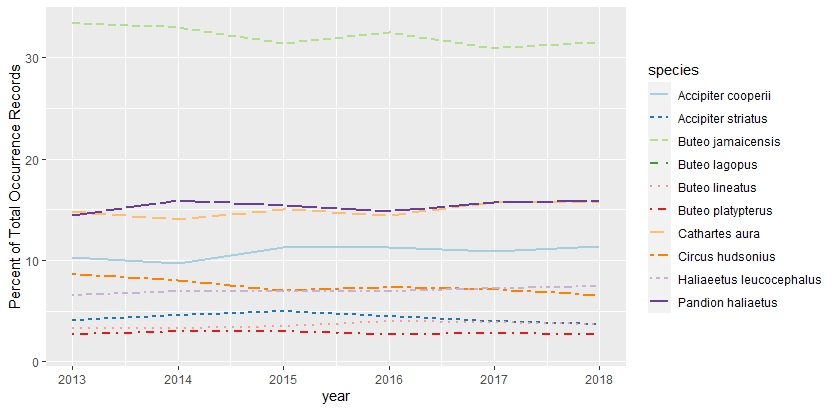


Table 2. Percent contributions of raptor species to total raptor occurrence records in 2018.

|  |  |  |
| --- | --- | --- |
| Species | Number of Occurrences | Percent of Occurrences in 2018 |
| *Buteo jamaicensis* | 17596 | 31.4 |
| *Pandion haliaetus* | 8830 | 15.8 |
| *Cathartes aura* | 8779 | 15.7 |
| *Accipiter cooperii* | 6348 | 11.3 |
| *Haliaeetus leucocephalus* | 4130 | 7.3 |
| *Circus hudsonius* | 3639 | 6.5 |
| *Accipiter striatus* | 2052 | 3.6 |
| *Buteo lineatus* | 2041 | 3.6 |
| *Buteo platypterus* | 1517 | 2.7 |
| *Coragyps atratus* | 490 | 0.8 |
| *Buteo lagopus* | 432 | 0.7 |
| *Aquila chrysaetos* | 9 | 0.01 |
| *Accipiter gentilis* | 9 | 0.016107 |
| *Ictinia mississippiensis* | 5 | 0.008948 |

Figure 3. Percent contribution of each raptor species to the total number of occurrence records each year. Red-tailed hawk (*Buteo jamaicensis*), Osprey (*Pandion haliaetus*), and Turkey vulture (*Cathartes aura*) make up a majority of the records across the years (over 60%). Species that contributed less than 1% of records each year were not included in the figure (they include Black vulture, Rough-legged buzzard, Northern goshawk, Golden eagle, Mississippi Kite, and Swallow-tailed Kite).

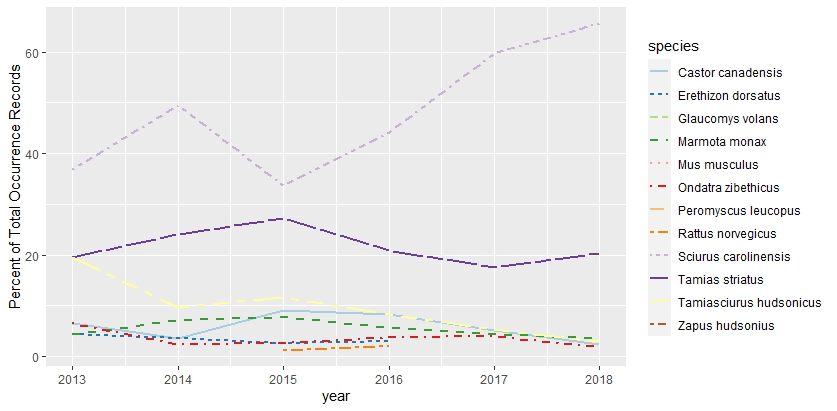


Figure 4. Percent contribution of each rodent species to the total number of occurrence records each year. Eastern gray squirrel (*Sciurus carolinensis*) and Eastern chipmunk (*Tamias striatus*) make up a majority of the records across the years (over 60%). The Groundhog (*Marmota monax*), American red squirrel (*Tamiasciurus hudsonicus*), and American beaver (Castor canadensis) are the next highest contributors. Rat and mice species which are large contributors to outages are not as well represented in the data.

Table 3. Number of records of species occurrences and outages for each year. There is a unanimous increase in the record availability from 2013-2018.

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Raptor Occurrences | Outage Occurrences | Rodent Occurrences |
| 2013 | 27346 | 22147 | 46 |
| 2014 | 32749 | 9511 | 83 |
| 2015 | 33736 | 18739 | 77 |
| 2016 | 42698 | 17340 | 206 |
| 2017 | 45106 | 27521 | 409 |
| 2018 | 55878 | 42895 | 1312 |

Figure 5. Number of records of species occurrences and outages for each year. There is a unanimous increase in the record availability from 2013-2018.

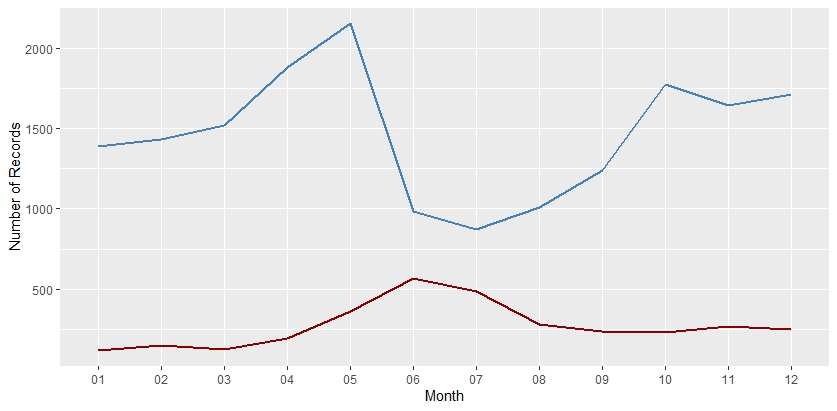


Figure 6. Distribution of Red-tailed hawk occurrences (blue) and animal caused outages (red) by months in 2018.

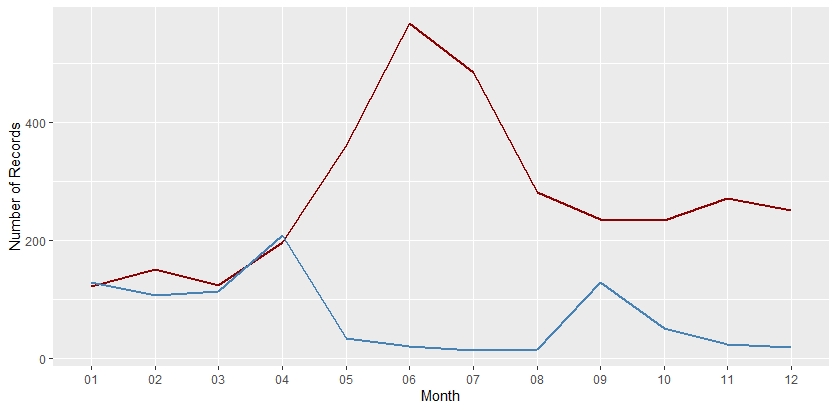


Figure 7. Distribution of Eastern gray squirrel occurrences (blue) and animal caused outages (red) by months in 2018.

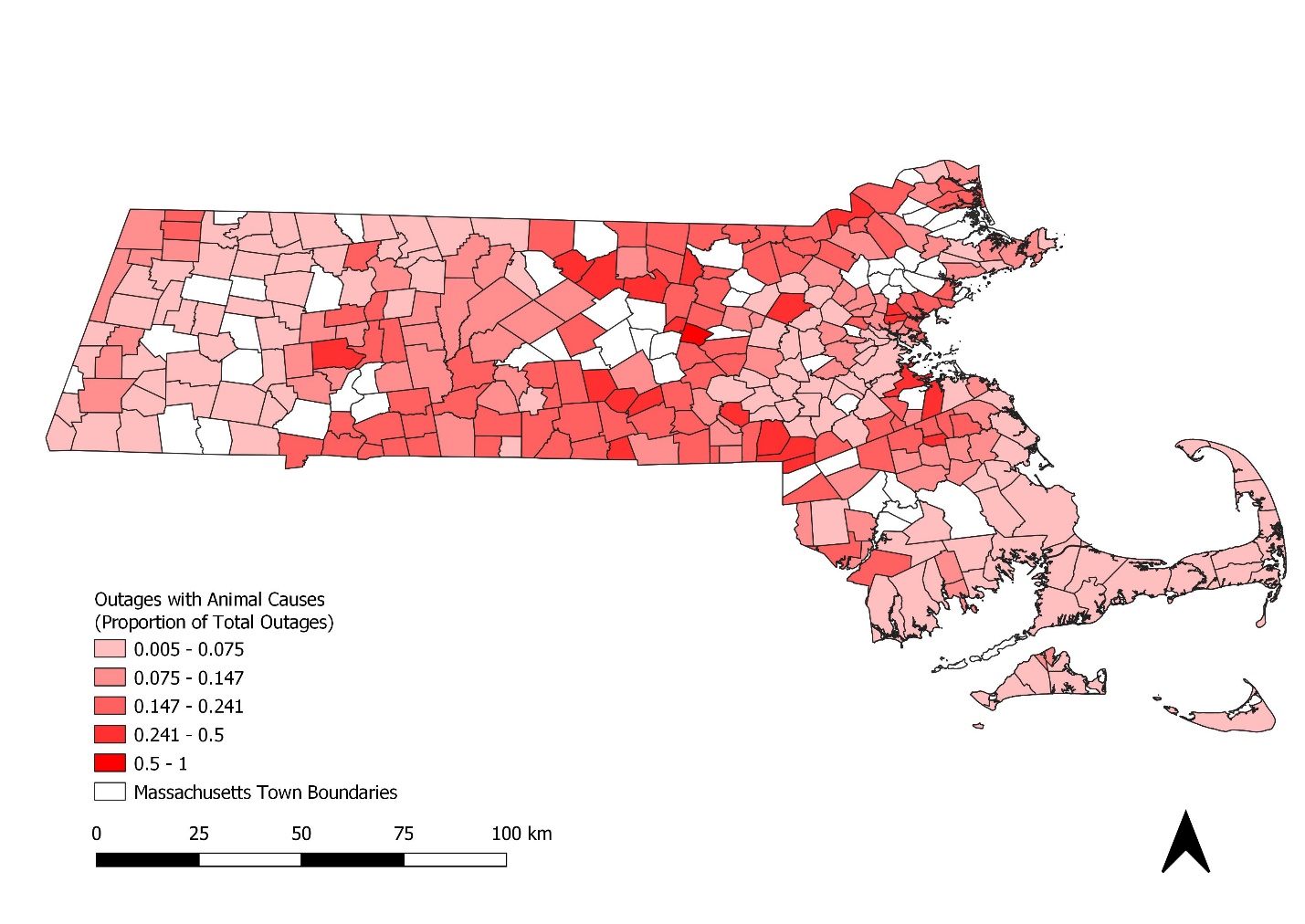
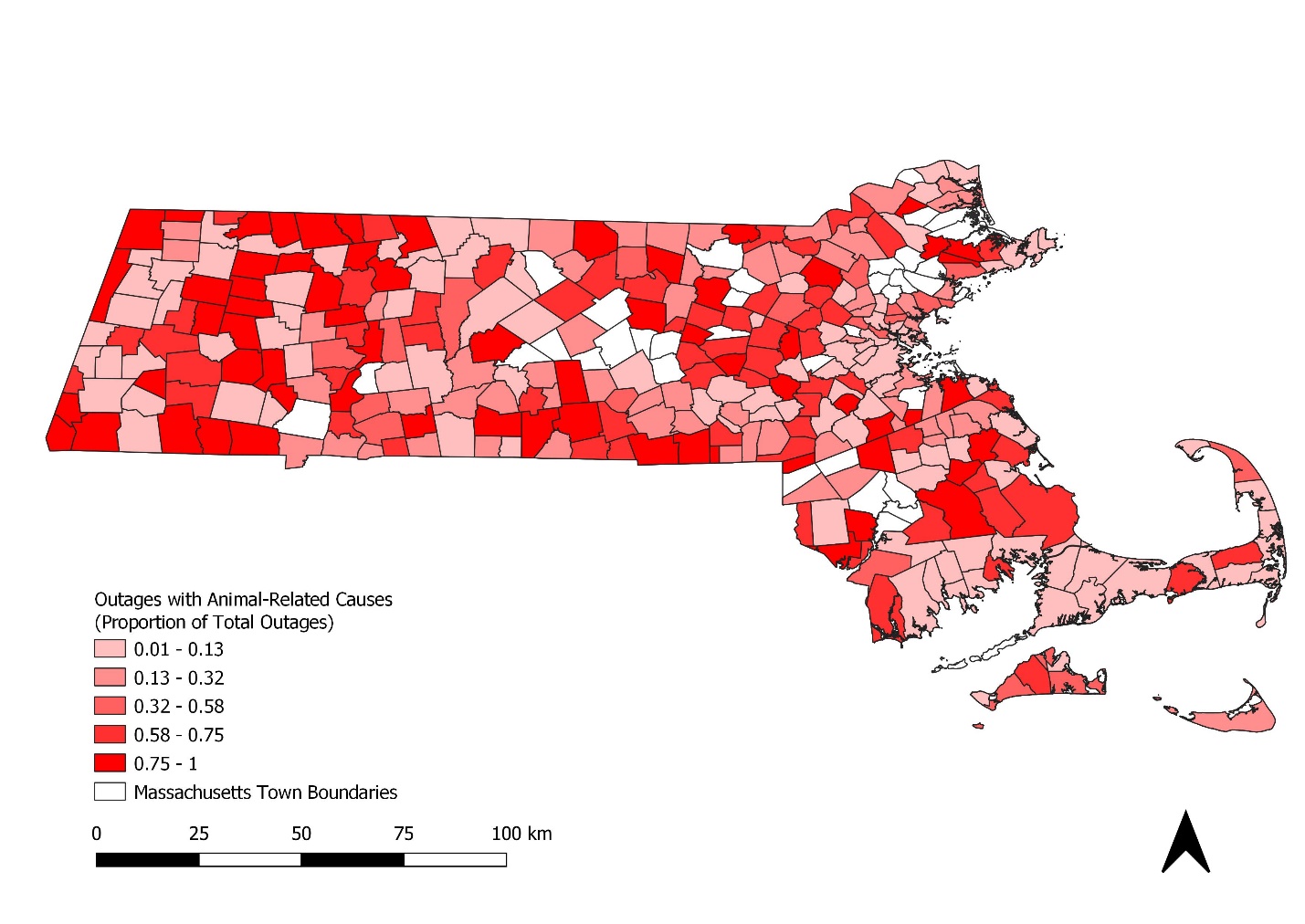
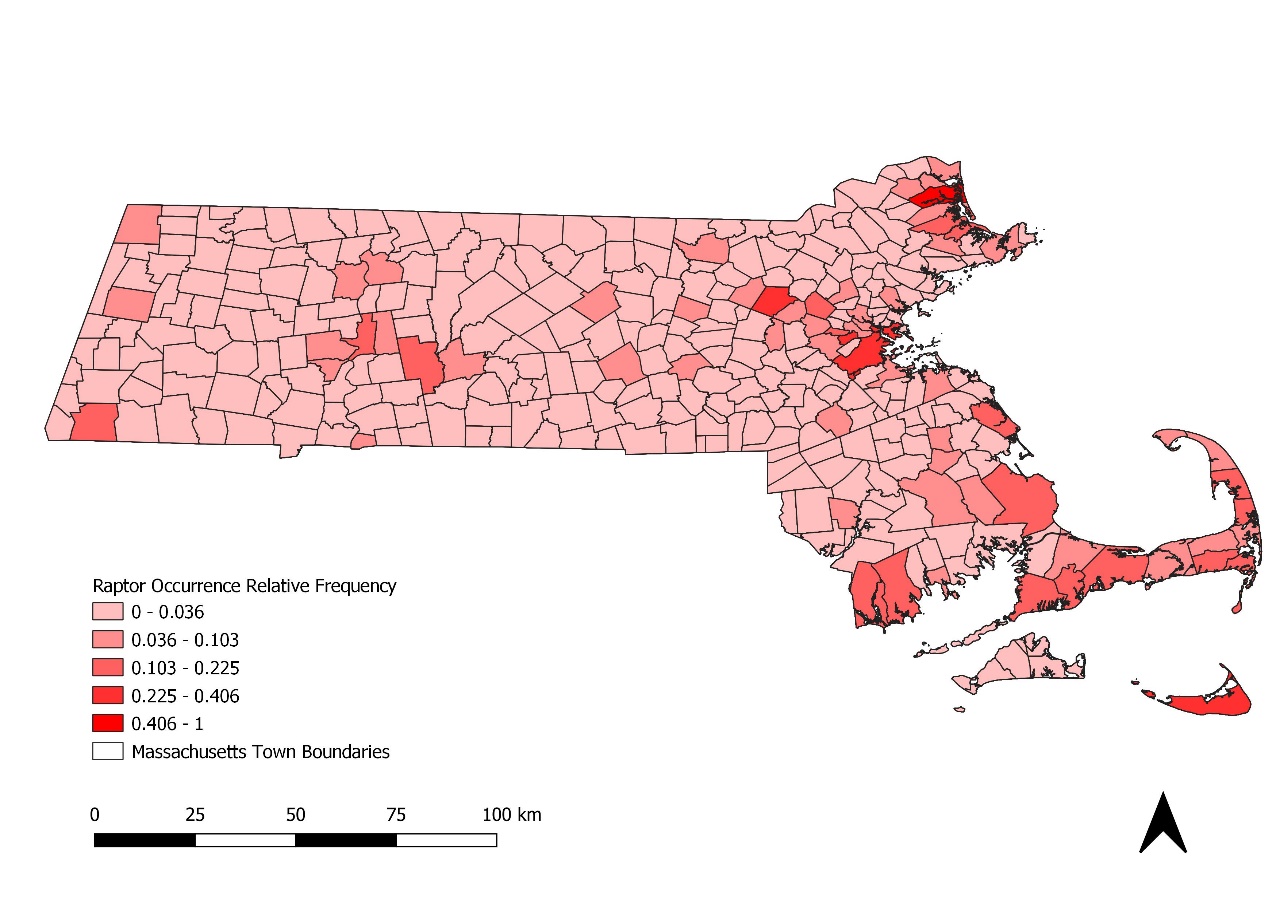
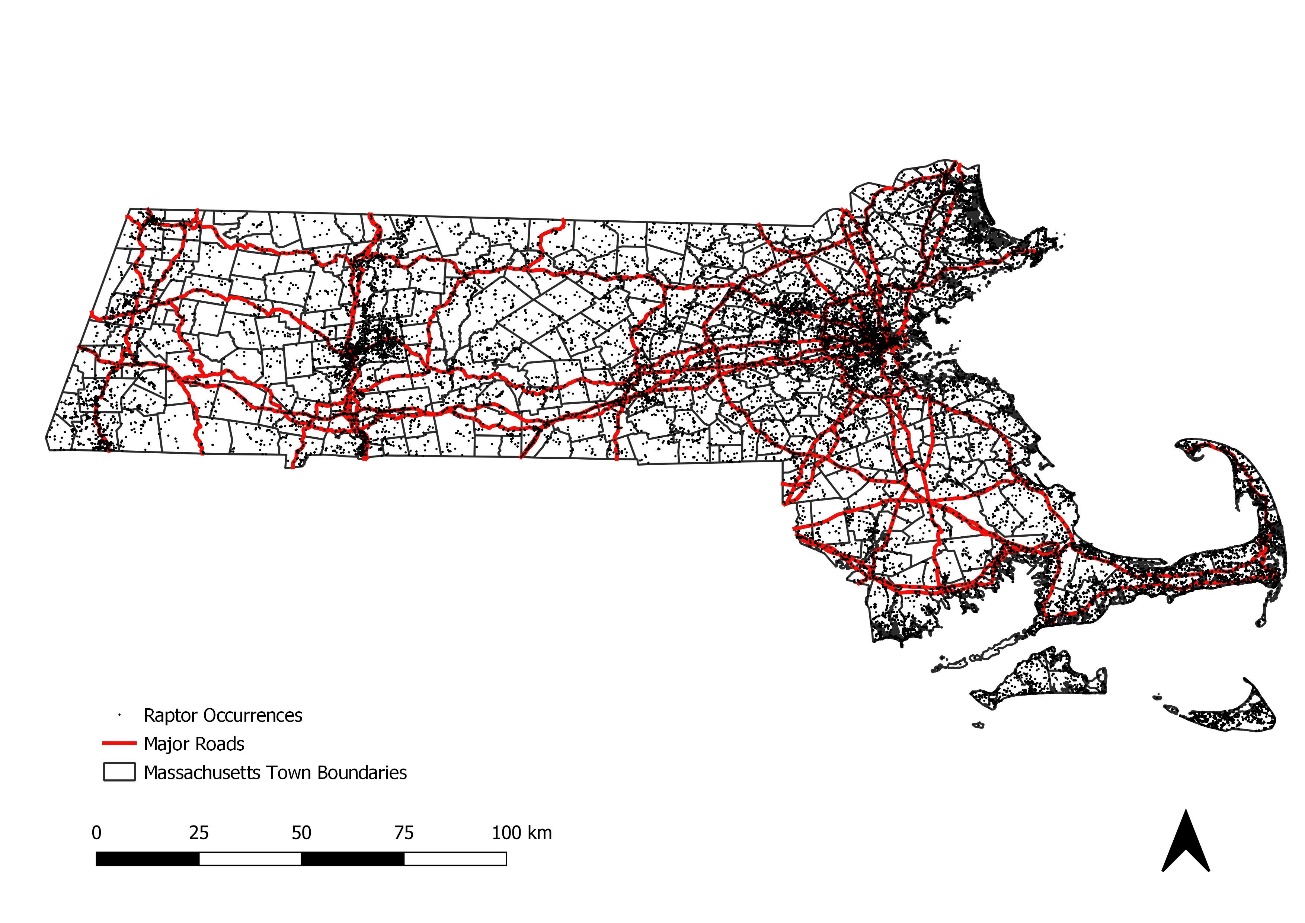


Figure 8. Proportion of total outages from 2013-2018 per town that had animal-related outage causes (Tree contact, Animal, Unknown, Physical Interference/Other).

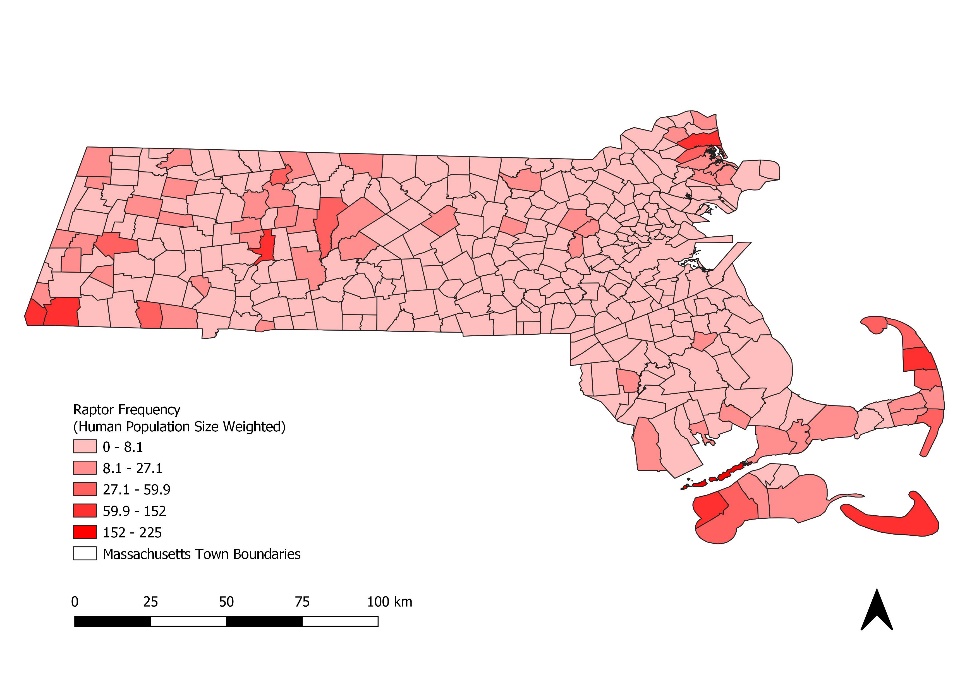
\*Relative outages, hot spots

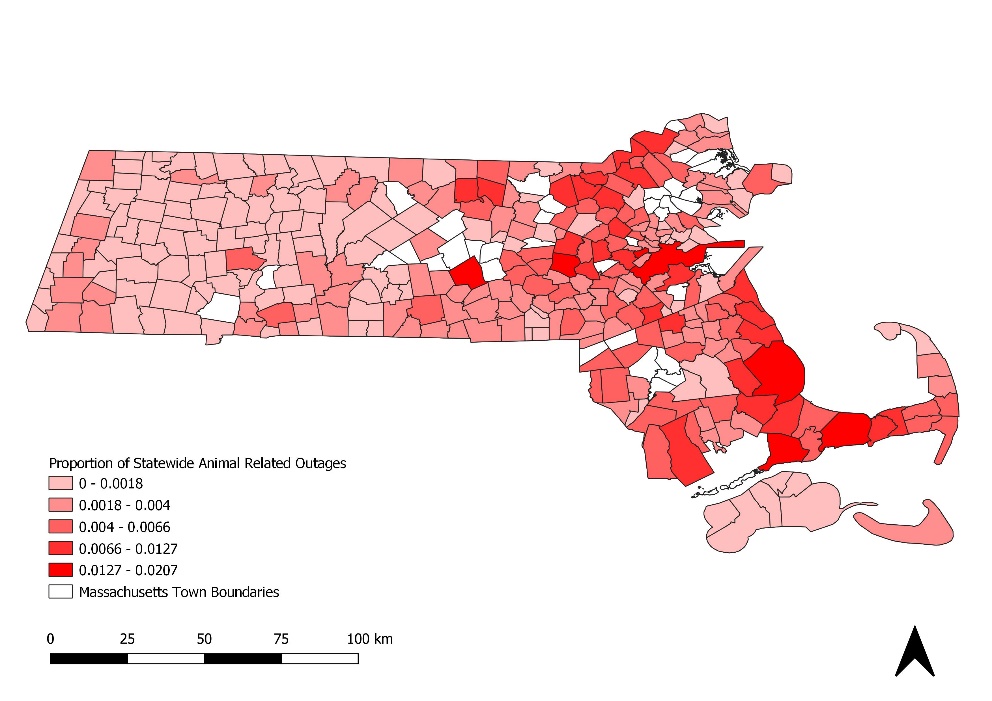
Figure 9. Proportion of total outages from 2013-2018 per town that had animal outage causes (Reason.For.Outage= ‘Animal’).





Figures 10 (top). Total raptor frequency from 2013-2018 relative by towns (1 being the highest overall frequency). Figure 11 (bottom). Raptor occurrence locations overlaid on major roads. Occurrences appear spatially correlated to urban locations and roads where detection probabilities increase.





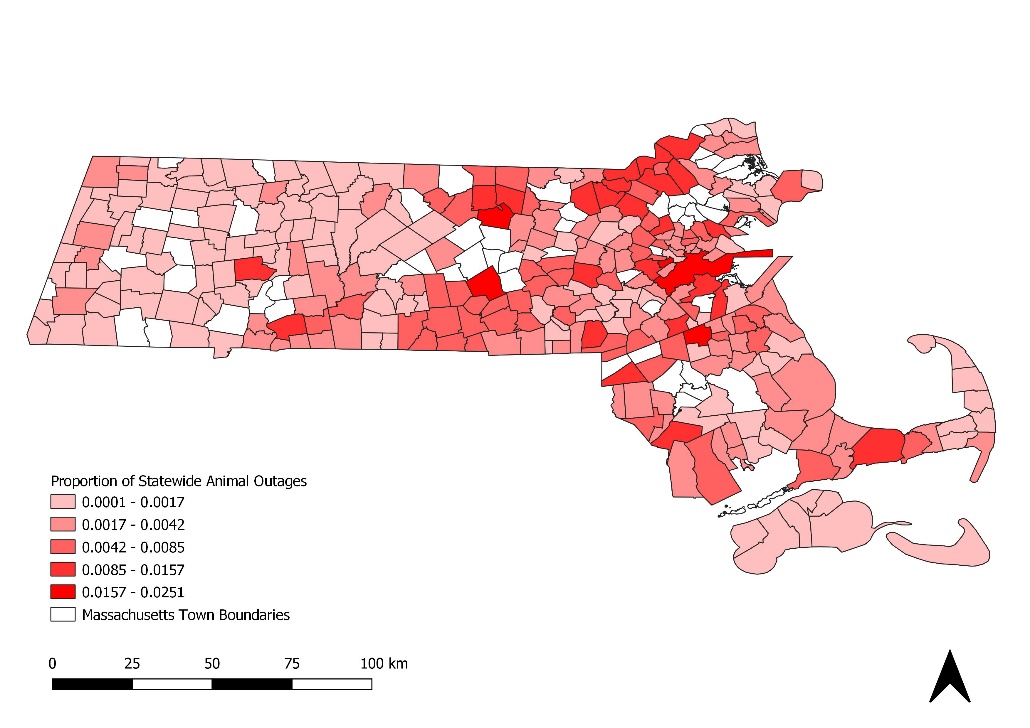


Figure 12 (top). Total raptor occurrence frequency by town (2013-2018) weighted by the town’s human population size in 2010. Animal and Animal related outage maps are below for spatial reference.

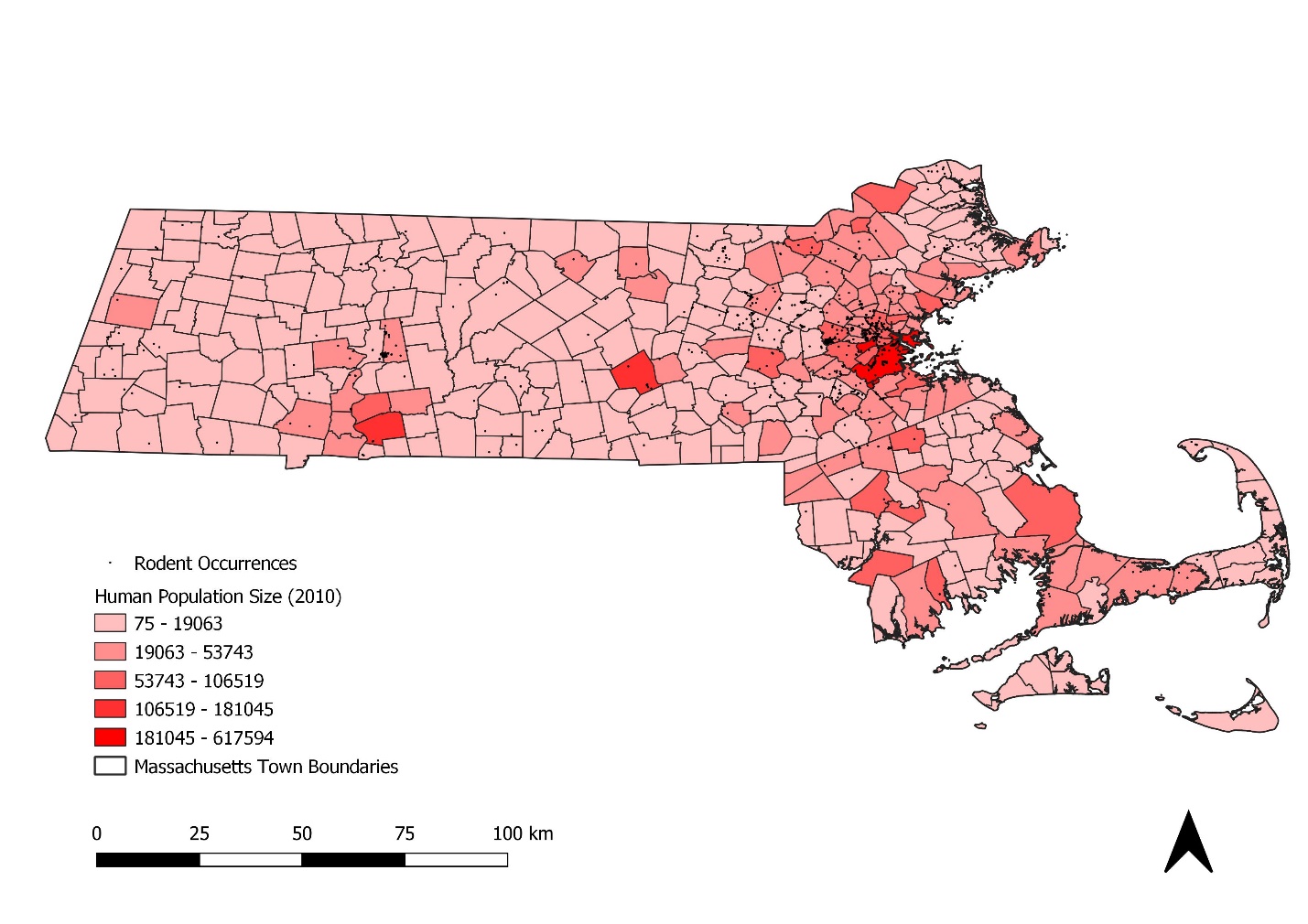


Figure 13. Rodent occurrence locations overlaid on town human population sizes in 2010. Occurrences appear spatially correlated to urban locations where detection probabilities increase.

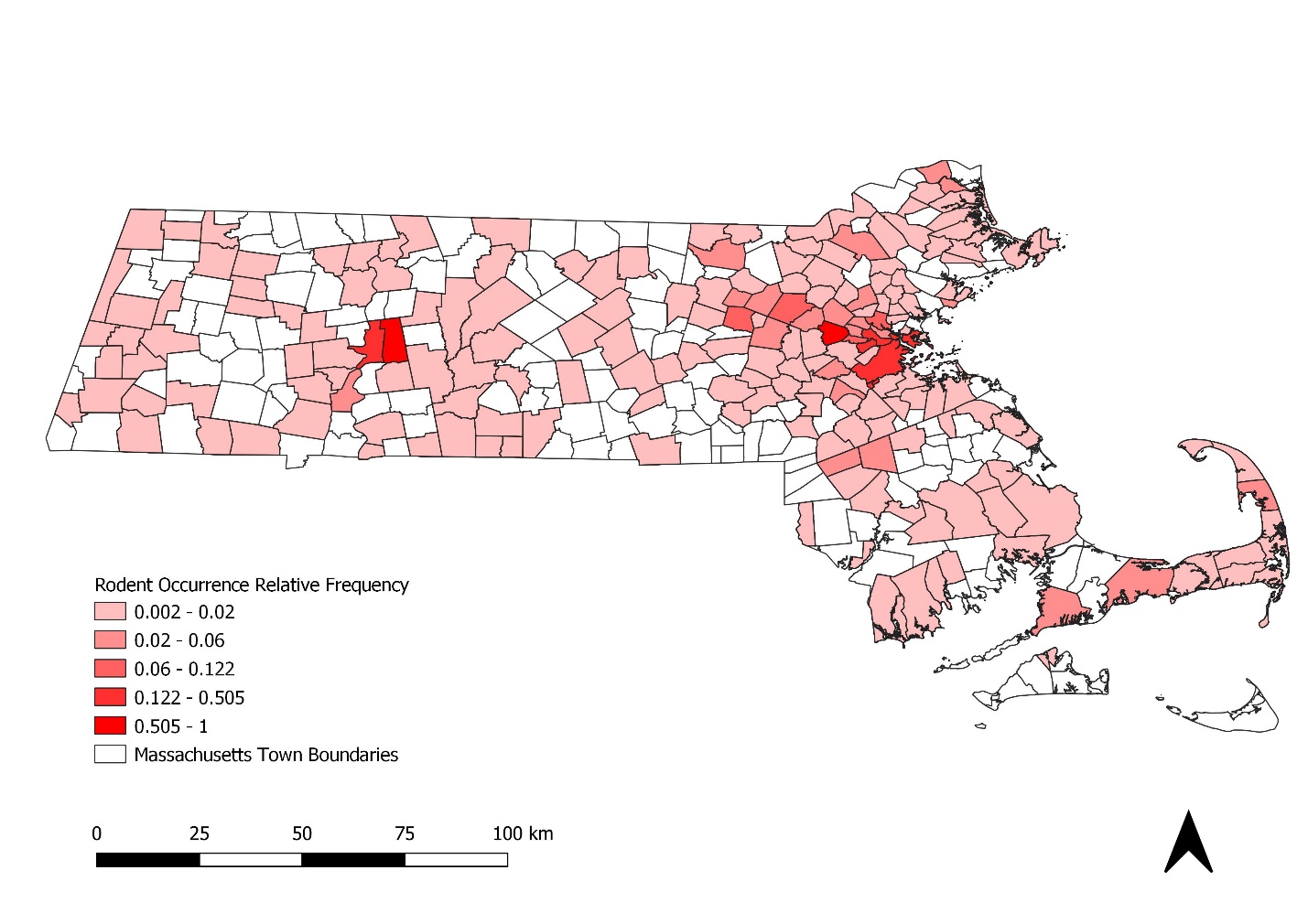
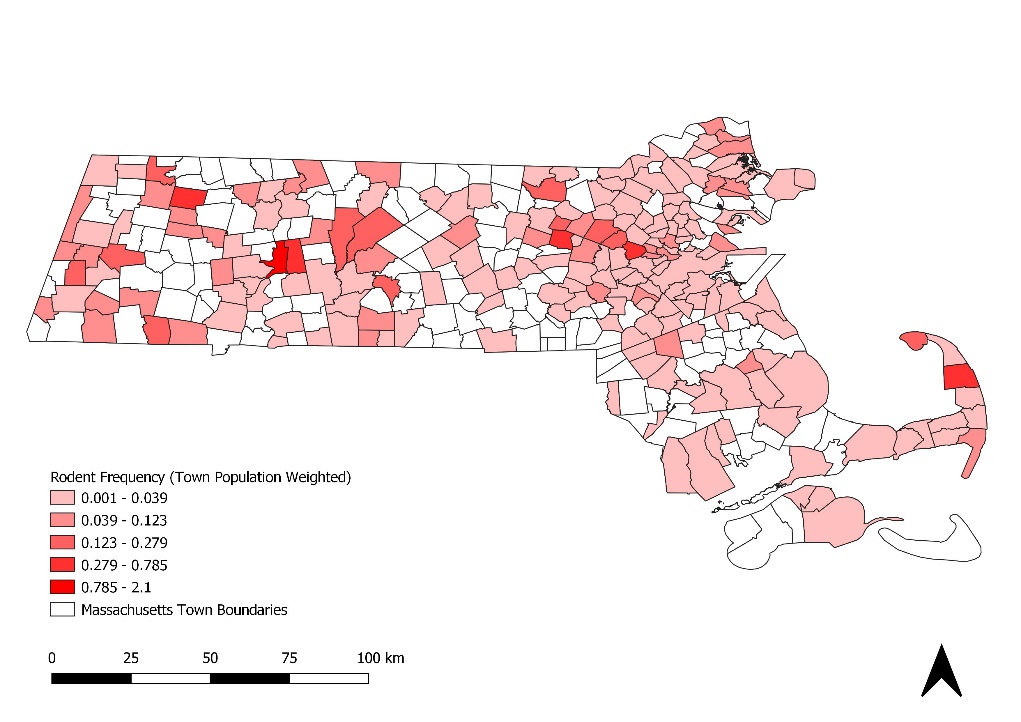
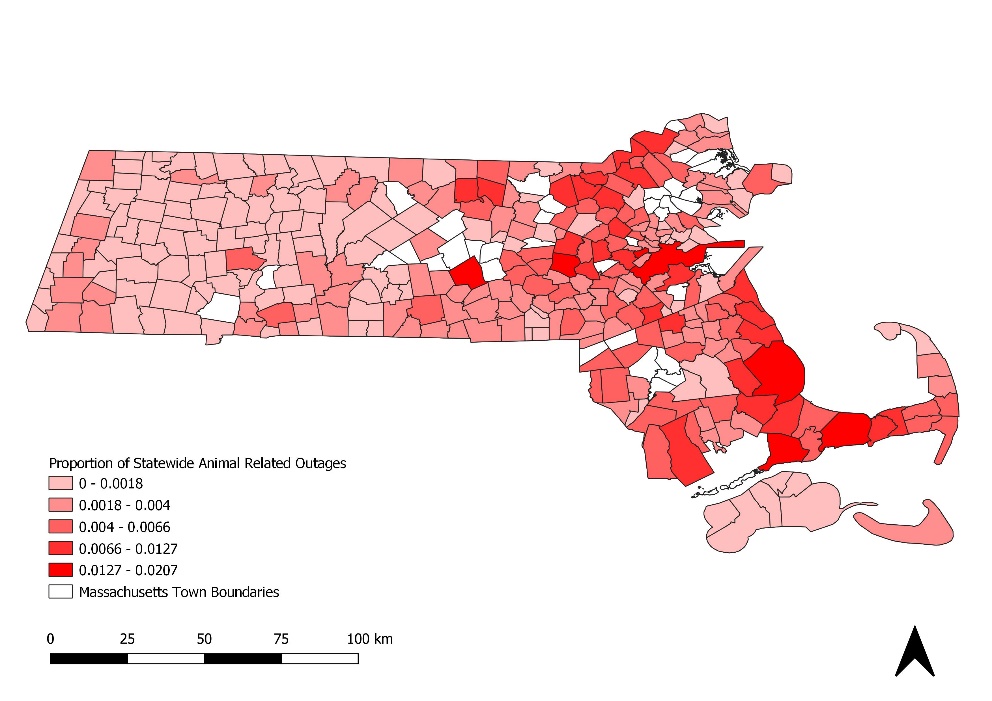


Figure 14. Total rodent frequency from 2013-2018 relative by towns (1 being the highest overall frequency).





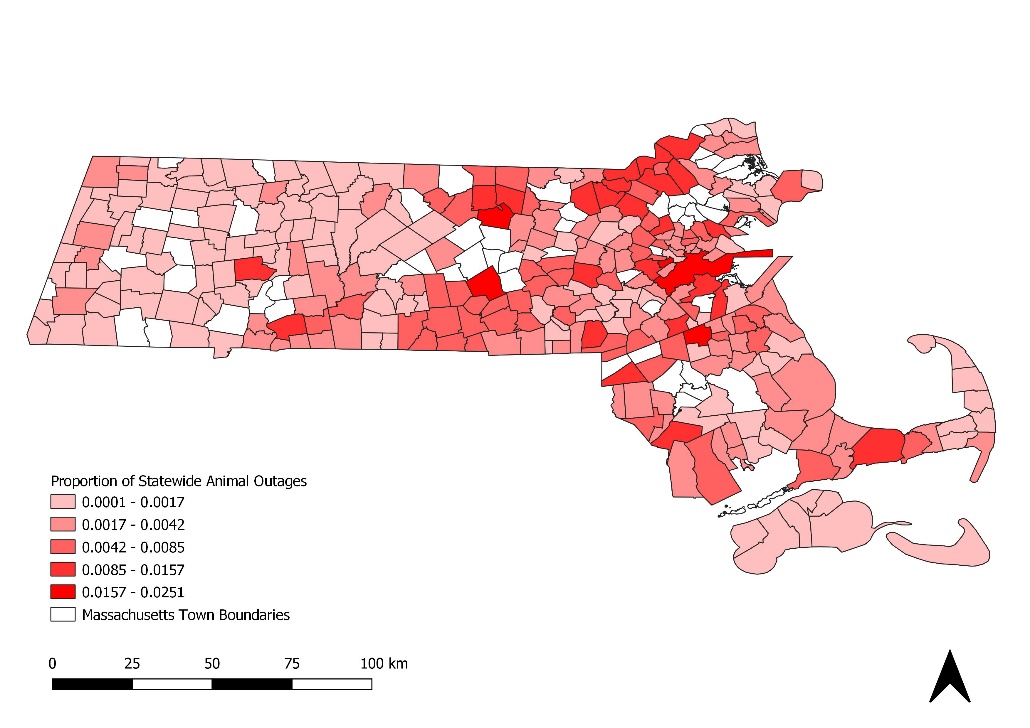
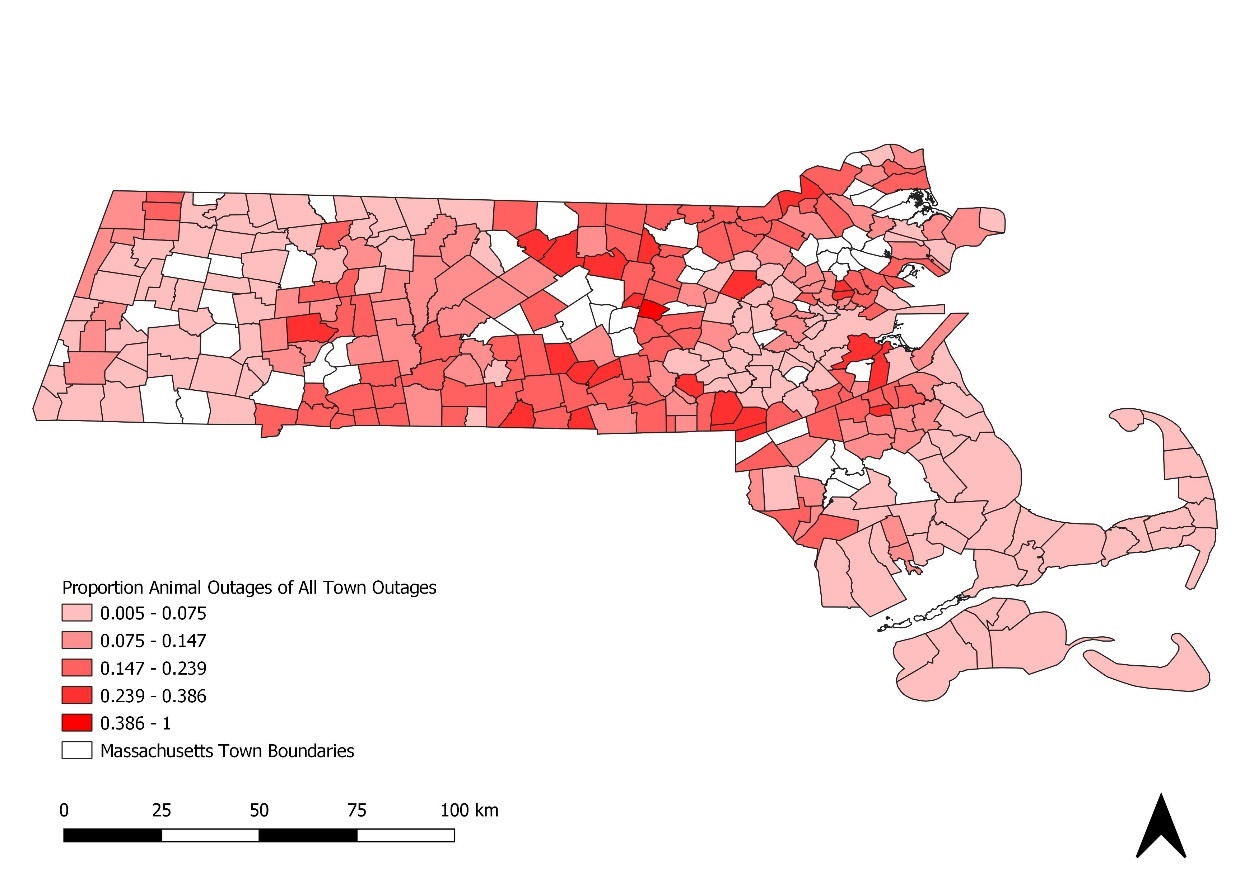


Figure 15 (Top). Total rodent occurrence frequency by town (2013-2018) weighted by the town’s human population size in 2010. There are areas of overlap with animal-related and animal outages (below).



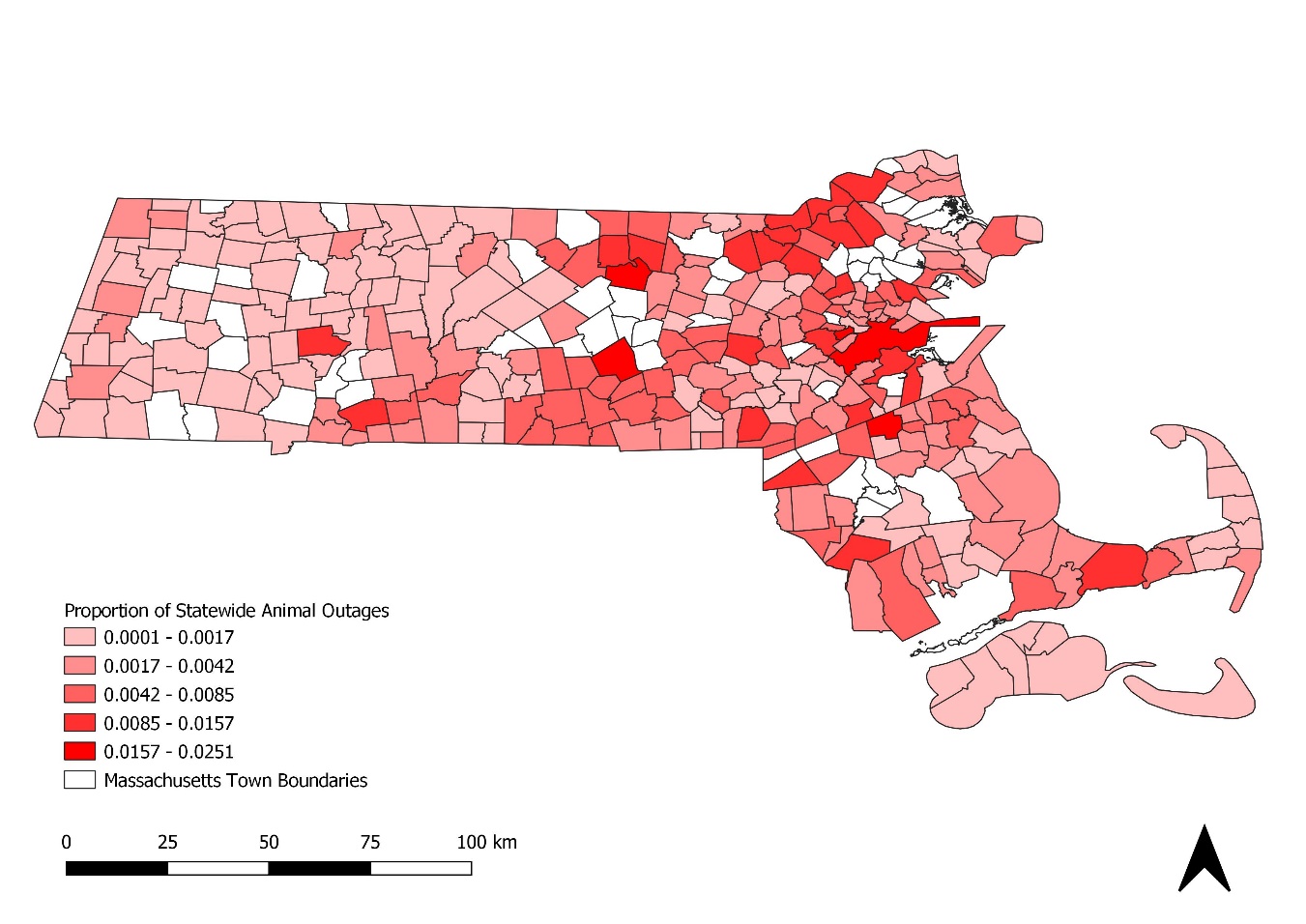
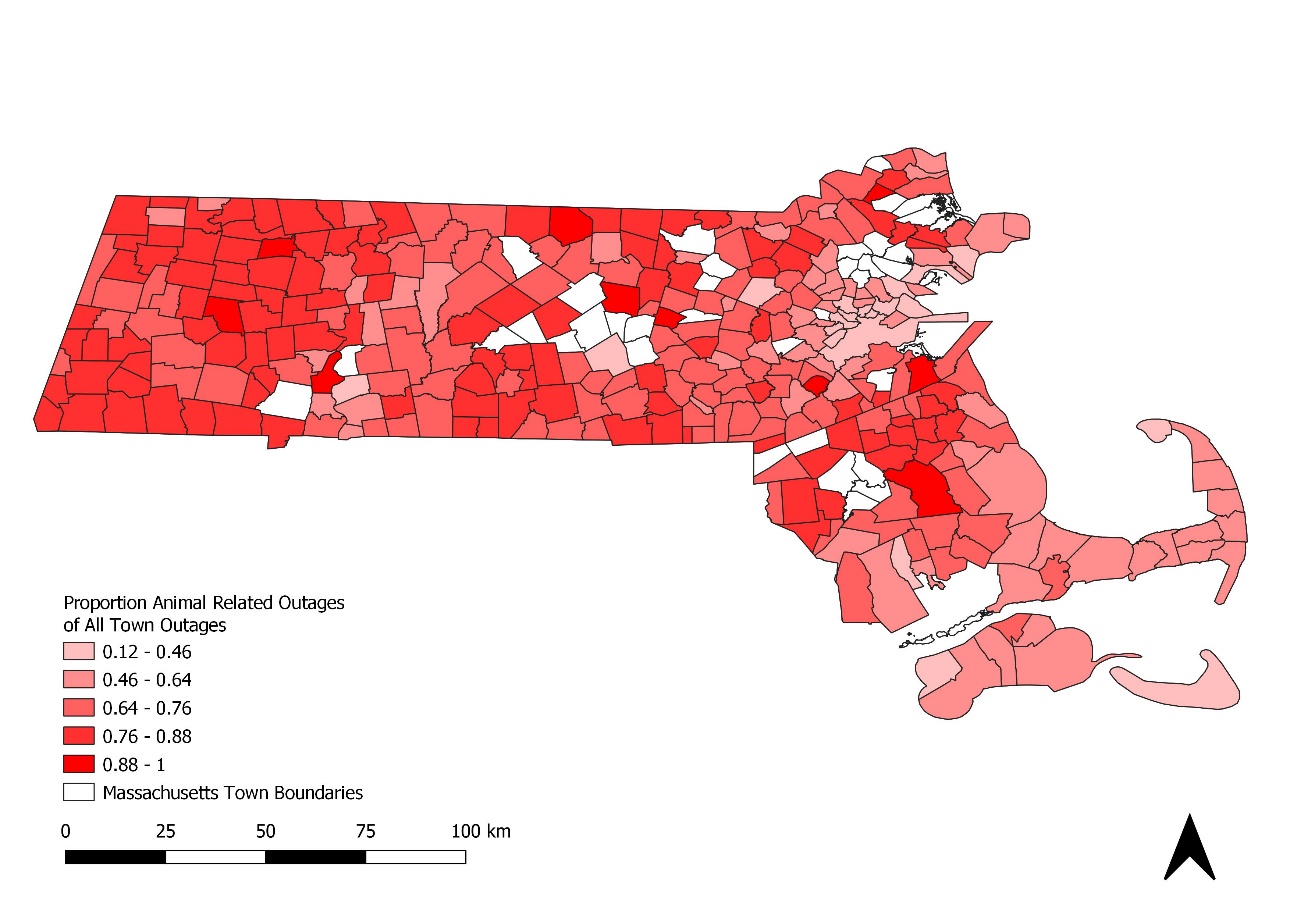


Figure. Comparing two measures of animal outage reports. (Top) The proportion of each town’s outages that are animal caused. (Bottom) The proportion of the total number of animal caused outages statewide contained in each town.



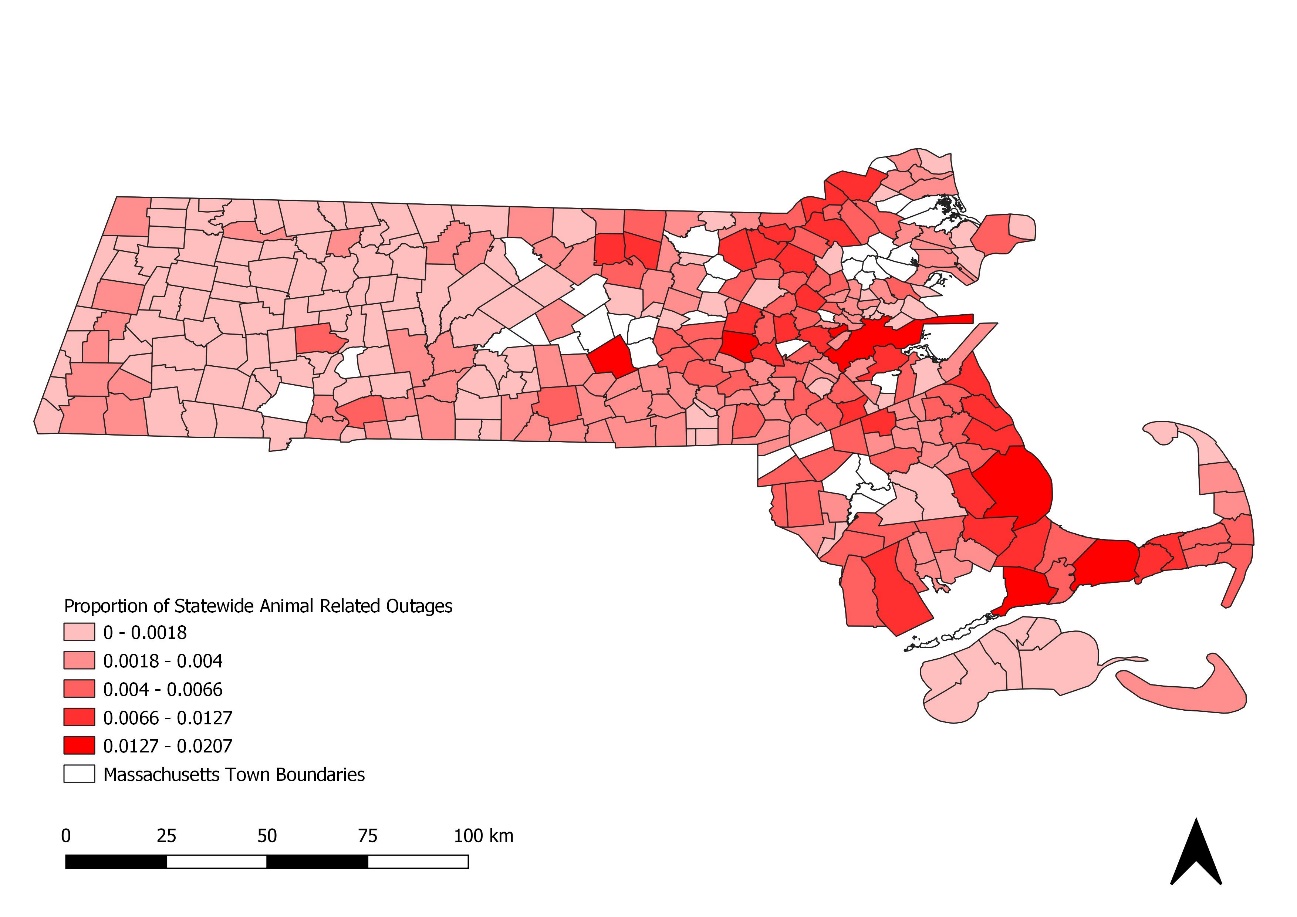
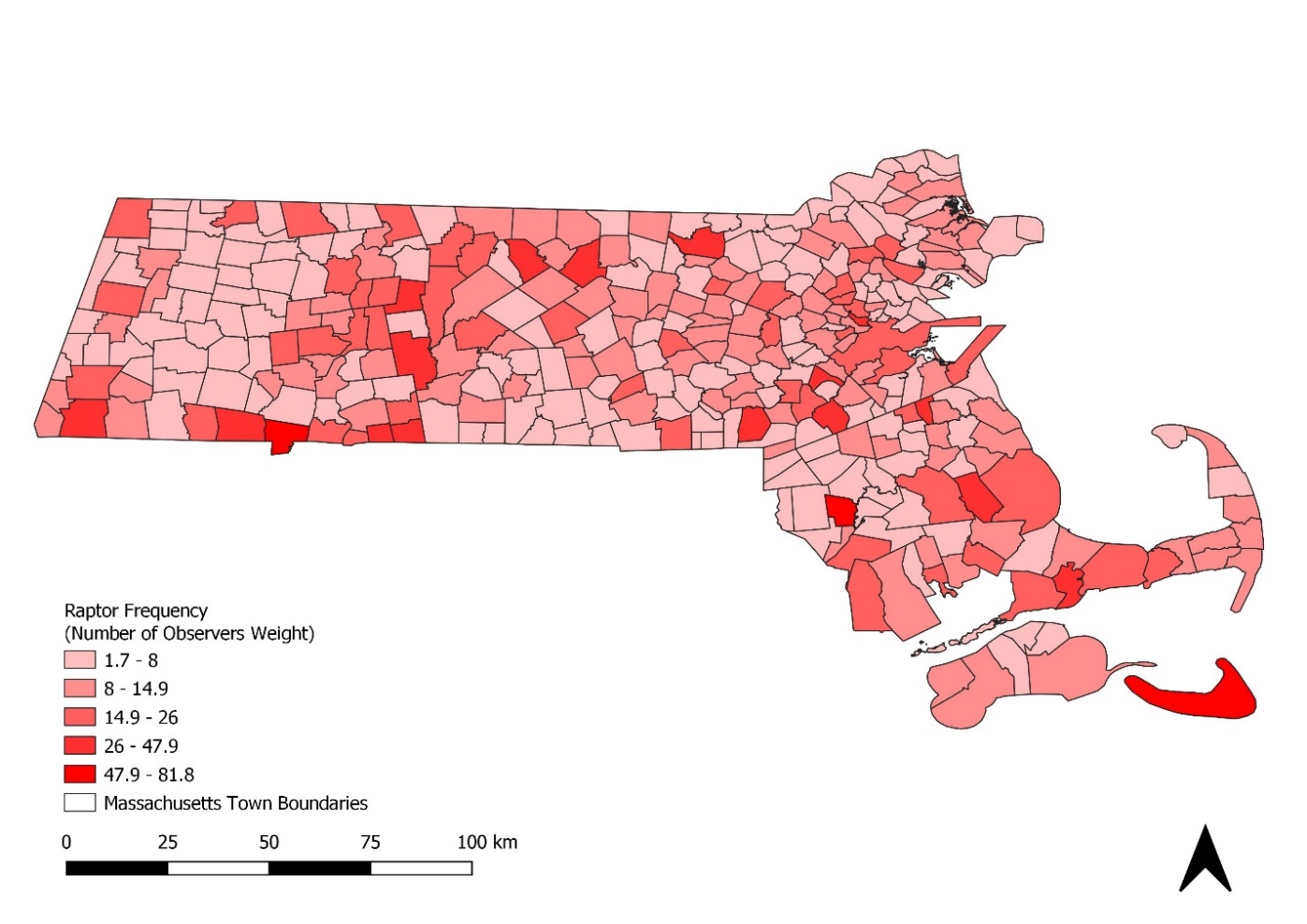


Figure. Comparing two measures of animal related outage reports. (Top) The proportion of each town’s outages that are from animal related causes. (Bottom) The proportion of the total number of outages statewide contained in each town that have animal related causes.



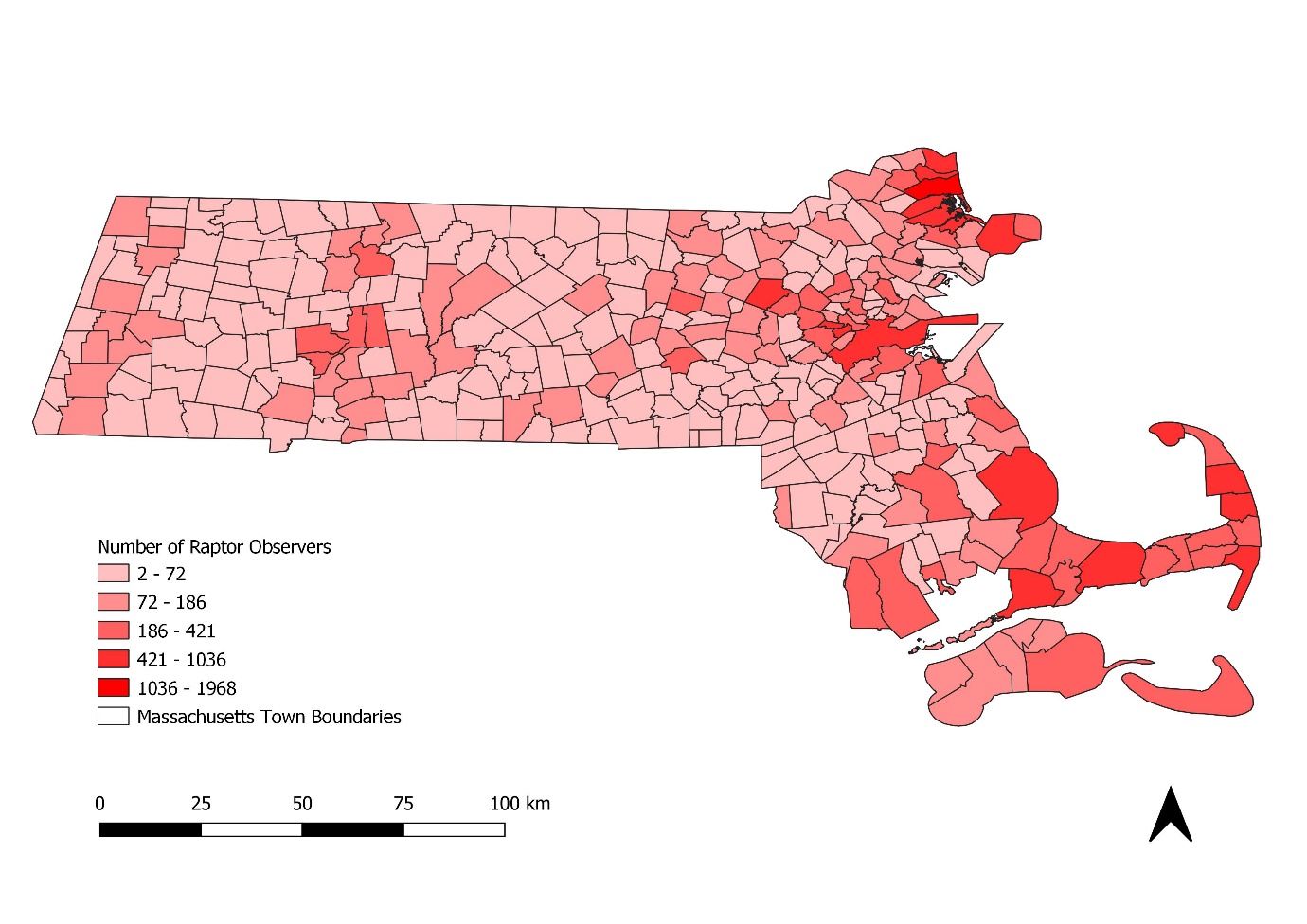
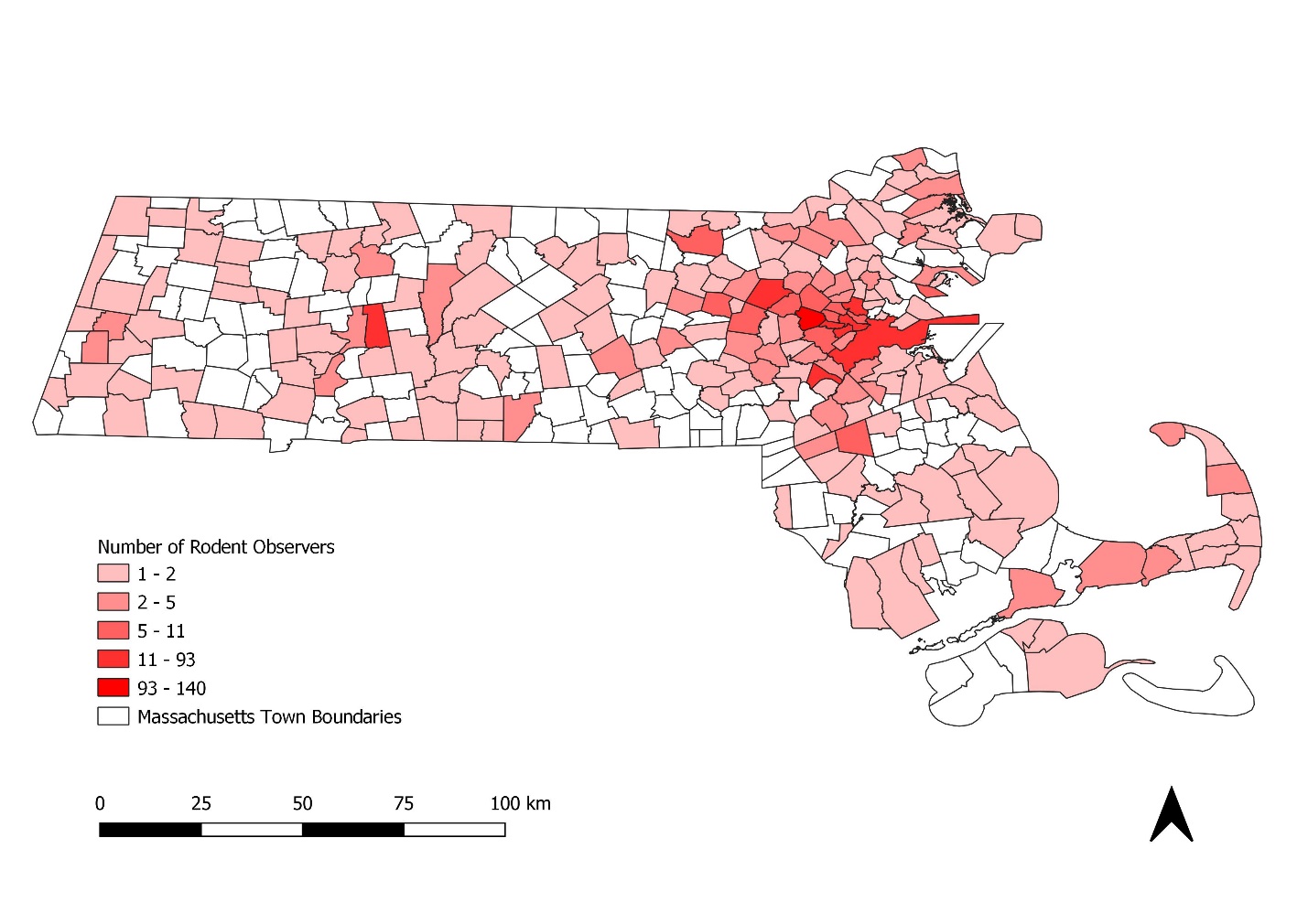


Figure . (Top) Raptor frequency totaled from 2013-2018 and weighted by the number of observers per town. (Bottom) Total number of unique raptor observers per town.



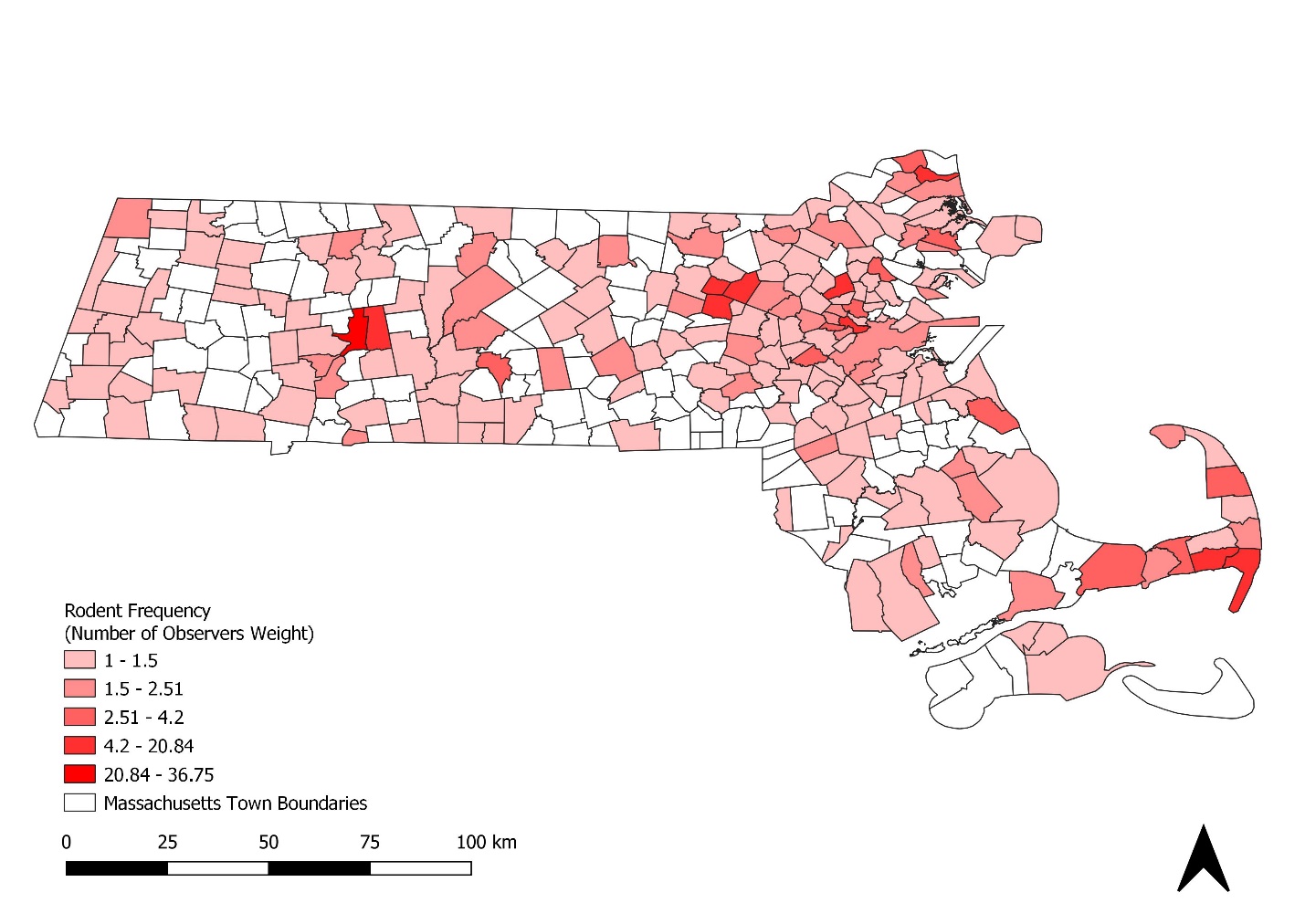


Figure . (Top) Rodent frequency totaled from 2013-2018 and weighted by the number of observers per town. (Bottom) Total number of unique rodent observers per town.

Time Series Tables:

Temporal scales: annual, monthly, daily, and time of day (simplified into day and night classifications)

Day and night classifications were only applied to the rodent and outage data since raptor occurrences have minimal records with time stamp temporal resolution. The classifications were determined by dividing months into “winter” months (October-March) and “summer” months (April-September) and applying summer solstice sunrise and sunset times to summer months and winter solstice times to winter months. Occurrence and outage times that were equal to or within the sunrise/sunset times were counted as occurring during the day and given a 1 in a new day.night field. Times that were greater than the sunset time or less than the sunrise time were given a 0 in the day.night field and considered nightly events.

Spatial scales: Towns

Summary fields:

* For outages
  + reason.count- the number of outages within the town during the specified time scale caused by a specified outage reason category
  + all.count- the total number of outages within the town during the specified time scale (all outage reasons)
  + norm.outs.day- the proportion of reason.count / all.count, number of outages in the town during the time scale from a specified cause out of the total outages in the town during the time scale.
  + reason.count.state-The number of outages with the specified outage reason across the state during the specific time scale.
  + prop.state.outs- The proportion of reason.count / reason.count.state, proportion of outages from a specific cause in the town out of the total number of outages from that cause across the state during the specified time scale.
* For species occurrences
  + freq.perc- number of occurrences per town out of the number of occurrences statewide during the specified time scale.
  + freq.rel - number of occurrences per town relative to the highest number of town occurrences during the specified time scale (number of occurrences at town x / highest number of town occurrences across all towns).
  + freq.weighted- number of occurrences per town during the specified time scale weighted by the town human population size.
  + freq.obs.weight- number of occurrences per town weighted by the number of observers in the town.
  + Density- number of occurrences per town during the specified time scale divided by the area of the town.

Summary figures:

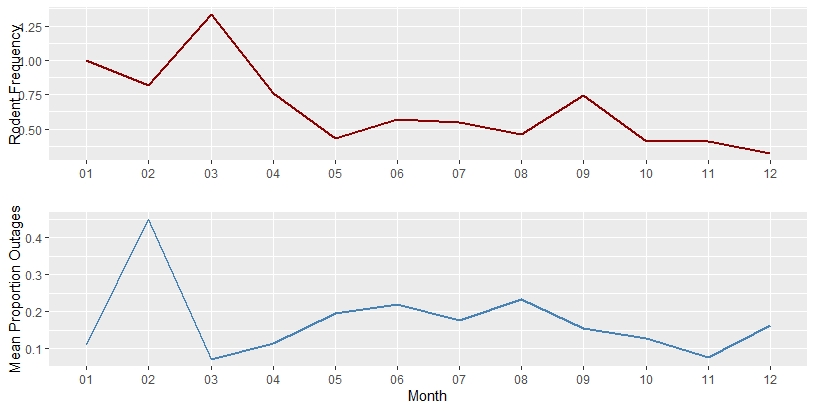


Figure . Monthly rodent frequency weighted by number of observers (top) and the proportion of squirrel caused outages out of total outages (bottom) averaged across towns statewide during 2018.

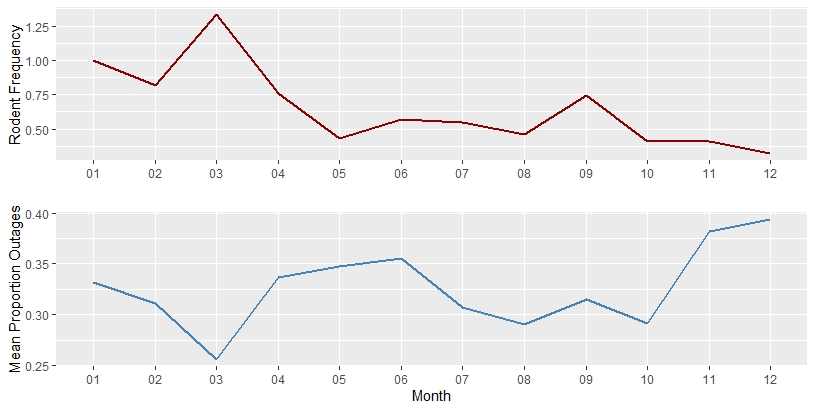


Figure . Monthly rodent frequency weighted by number of observers (top) and the proportion of animal caused outages out of total outages (bottom) averaged across towns statewide during 2018.

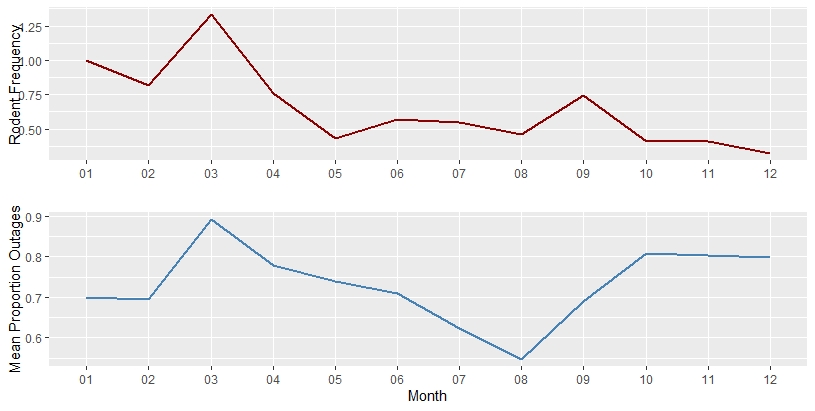
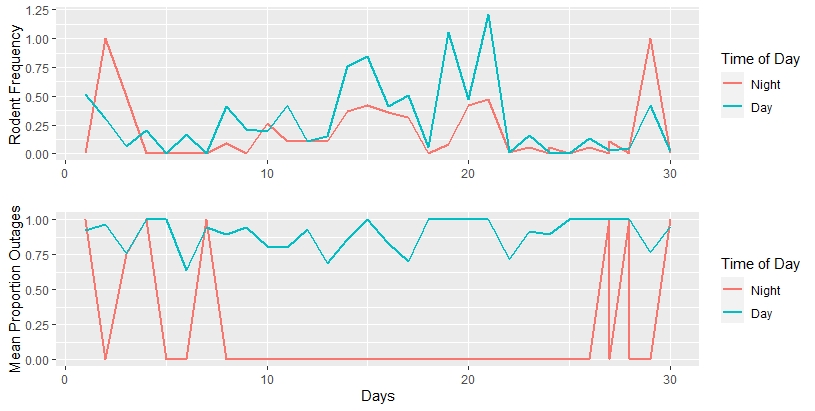


Figure . Monthly rodent frequency weighted by number of observers (top) and the proportion of outages with animal related causes out of total outages (bottom) averaged across towns statewide during 2018.



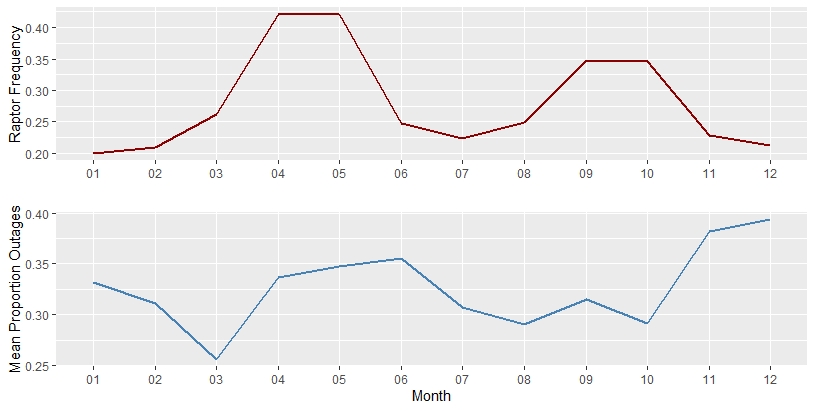


Figure . Monthly raptor frequency weighted by number of observers (top) and the proportion of animal caused outages out of total outages (bottom) averaged across towns statewide during 2018.

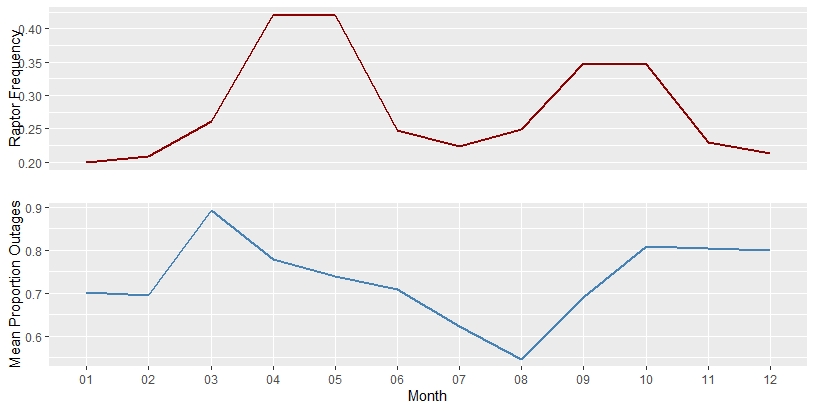


Figure . Monthly rodent frequency weighted by number of observers (top) and the proportion of outages with animal related causes out of total outages (bottom) averaged across towns statewide during 2018.