

Modern R with tidyverse [Solutions]

Jonathan de Bruin, Barbara Vreede

Contents

1. Read and save data	2
Basic exercise I - Read data into R	2
a) Read the data <code>HUJI_Crane_Israel_GPRS.csv</code> into R	2
b) Load <code>readxl</code> (to read Excel files)	3
c) Read the additional observations <code>crane_additional_observations.xlsx</code> into R.	3
Basic exercise II - Dataset properties	3
Optional exercise (+) - Save data to a CSV file with delimiter ;.	4
Optional exercise (++) - Write tibble <code>data_crane</code> to an Excel file.	4
Optional exercise (++) - Read and write SPSS, SAS, and STATA data files	5
Optional exercise (+++) - Parse datetime columns	7
2. Data visualisation	9
Basic exercise I - Quick plots of the <code>data_crane</code>	9
a) Single column plots	9
b) Two column plots	13
Basic exercise II - Using ggplot for plotting	17
Optional exercise (+) - Statistical layers for graphs.	20
Optional exercise (+) - Scale axes	21
Optional exercise (++) - Plot the crane positions on a map	22
a) Install the package <code>maps</code>	22
b) Plot the crane data on a map.	22
c) Use an individual identifier to colour the different cranes.	23
Optional exercise (+++) - Create facets.	24
3. Data transformation	26
Basic exercise I - Subset data	26
a) Filter data from one individual crane	26
b) Filter data with complete GPS information	26
c) Select specific columns from your data	27
d) Combine <code>filter</code> and <code>select</code> using the pipe operator	27
Basic exercise II - Compute the magnitude of the magnetic field	27
Optional exercise (+) - Summarise results	28
Optional exercise (++) - Join datasets	29
Optional exercise (+++) - gather all acceleration data	29

*This document is part of the workshop **Introduction to R & Data** by Utrecht University RDM Support.*

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.1.0      v purrr   0.3.0
## v tibble  2.0.1      v dplyr   0.8.0.1
## v tidyrr  0.8.2      v stringr 1.4.0
## v readr   1.3.1      vforcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

1. Read and save data

Basic exercise I - Read data into R

a) Read the data HUJI_Crane_Israel_GPRS.csv into R

```
data_crane <- read_csv('data/HUJI_Crane_Israel_GPRS.csv')

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   visible = col_logical(),
##   timestamp = col_datetime(format = ""),
##   eobs_activity = col_logical(),
##   eobs_activity_samples = col_logical(),
##   eobs_start_timestamp = col_character(),
##   eobs_status = col_character(),
##   import_marked_outlier = col_logical(),
##   orn_transmission_protocol = col_character(),
##   sensor_type = col_character(),
##   individual_taxon_canonical_name = col_character(),
##   individual_local_identifier = col_character(),
##   study_name = col_character()
## )

## See spec(...) for full column specifications.

head(data_crane)

## # A tibble: 6 x 42
##   event_id visible timestamp           location_long location_lat
##       <dbl>    <lgl>    <dttm>              <dbl>        <dbl>
## 1 3.79e9  TRUE  2017-10-15 00:00:16      40.3        55.3
## 2 3.80e9  TRUE  2017-10-15 00:00:16      40.3        55.3
## 3 3.79e9  TRUE  2017-10-15 02:00:23      40.3        55.3
## 4 3.80e9  TRUE  2017-10-15 02:00:23      40.3        55.3
## 5 3.79e9  TRUE  2017-10-15 04:00:23      40.3        55.3
## 6 3.80e9  TRUE  2017-10-15 04:00:23      40.3        55.3
## # ... with 37 more variables: acceleration_raw_x <dbl>,
## #   acceleration_raw_y <dbl>, acceleration_raw_z <dbl>,
## #   bar_barometric_height <dbl>, battery_charge_percent <dbl>,
## #   battery_charging_current <dbl>, eobs_activity <lgl>,
## #   eobs_activity_samples <lgl>, eobs_battery_voltage <dbl>,
## #   eobs_fix_battery_voltage <dbl>,
## #   eobs_horizontal_accuracy_estimate <dbl>, eobs_key_bin_checksum <dbl>,
## #   eobs_speed_accuracy_estimate <dbl>, eobs_start_timestamp <chr>,
## #   eobs_status <chr>, eobs_temperature <dbl>, eobs_type_of_fix <dbl>,
## #   eobs_used_time_to_get_fix <dbl>, external_temperature <dbl>,
## #   gps_hdop <dbl>, gps_satellite_count <dbl>, gps_time_to_fix <dbl>,
## #   ground_speed <dbl>, heading <dbl>, height_above_ellipsoid <dbl>,
## #   height_above_msl <dbl>, import_marked_outlier <lgl>,
## #   gls_light_level <dbl>, mag_magnetic_field_raw_x <dbl>,
## #   mag_magnetic_field_raw_y <dbl>, mag_magnetic_field_raw_z <dbl>,
## #   orn_transmission_protocol <chr>, tag_voltage <dbl>, sensor_type <chr>,
```

```
## #   individual_taxon_canonical_name <chr>,
## #   individual_local_identifier <chr>, study_name <chr>
```

b) Load `readxl` (to read Excel files)

```
library(readxl)
```

c) Read the additional observations `crane_additional_observations.xlsx` into R.

```
data_crane_additional <- read_excel('data/crane_additional_observations.xlsx')
head(data_crane_additional)
```

```
## # A tibble: 6 x 4
##       event_id meas_1 meas_2 meas_3
##       <dbl>    <dbl>    <dbl>    <dbl>
## 1 3794510958    1.15    1.93    0.503
## 2 3796034354    0.376    0.307    0.341
## 3 3794510962    0.217    0.829    0.963
## 4 3796034355    0.684    0.250    0.129
## 5 3794510961    1.30     1.87    1.14
## 6 3796034356    0.847    5.14    0.964
```

Basic exercise II - Dataset properties

```
glimpse(data_crane)
```

```
## Observations: 20,873
## Variables: 42
## $ event_id                               <dbl> 3794510958, 3796034354, 3794...
## $ visible                                 <lgl> TRUE, TRUE, TRUE, TRUE, TRUE...
## $ timestamp                               <dttm> 2017-10-15 00:00:16, 2017-1...
## $ location_long                           <dbl> 40.31509, 40.31509, 40.31481...
## $ location_lat                            <dbl> 55.34764, 55.34764, 55.34792...
## $ acceleration_raw_x                     <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ acceleration_raw_y                     <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ acceleration_raw_z                     <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ bar_barometric_height                 <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ battery_charge_percent                <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ battery_charging_current              <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ eobs_activity                          <lgl> NA, NA, NA, NA, NA, NA, NA, ...
## $ eobs_activity_samples                  <lgl> NA, NA, NA, NA, NA, NA, NA, ...
## $ eobs_battery_voltage                  <dbl> 3784, 3818, 3784, 3818, 3784...
## $ eobs_fix_battery_voltage              <dbl> NA, 3801, NA, 3798, NA, 3796...
## $ eobs_horizontal_accuracy_estimate    <dbl> NA, 48.90, NA, 2.56, NA, 3.8...
## $ eobs_key_bin_checksum                 <dbl> 0, 2787211109, 0, 1067160123...
## $ eobs_speed_accuracy_estimate          <dbl> NA, 1.54, NA, 0.33, NA, 0.34...
## $ eobs_start_timestamp                  <chr> NA, "2017,15+Oct", NA, "2017...
## $ eobs_status                            <chr> NA, "A", NA, "A", NA, "A", N...
## $ eobs_temperature                      <dbl> NA, -7, NA, -7, NA, -8, NA, ...
## $ eobs_type_of_fix                       <dbl> NA, 3, NA, 3, NA, 3, NA, 3, ...
```

```

## $ eobs_used_time_to_get_fix
## $ external_temperature
## $ gps_hdop
## $ gps_satellite_count
## $ gps_time_to_fix
## $ ground_speed
## $ heading
## $ height_above_ellipsoid
## $ height_above_msl
## $ import_marked_outlier
## $ gls_light_level
## $ mag_magnetic_field_raw_x
## $ mag_magnetic_field_raw_y
## $ mag_magnetic_field_raw_z
## $ orn_transmission_protocol
## $ tag_voltage
## $ sensor_type
## $ individual_taxon_canonical_name
## $ individual_local_identifier
## $ study_name

glimpse(data_crane_additional)

## Observations: 20,873
## Variables: 4
## $ event_id <dbl> 3794510958, 3796034354, 3794510962, 3796034355, 37945...
## $ meas_1   <dbl> 1.1485201, 0.3756848, 0.2170456, 0.6836040, 1.2951363...
## $ meas_2   <dbl> 1.9269063, 0.3073029, 0.8286311, 0.2501690, 1.8683569...
## $ meas_3   <dbl> 0.50257465, 0.34084276, 0.96295022, 0.12898235, 1.136...

```

Optional exercise (+) - Save data to a CSV file with delimiter ;.

```

# create a directory for the output file
if (!dir.exists('tmp')){
  dir.create("tmp")
}

write_delim(data_crane, 'tmp/data_crane_csv_file.csv', delim = ';')

```

Optional exercise (++) - Write tibble data_crane to an Excel file.

```

install.packages("writexl")

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
library(writexl)

# create a directory
if (!dir.exists('tmp')){
  dir.create("tmp")
}

```

```

write_xlsx(data_crane, "tmp/data_crane.xlsx")

# or, if you want to write to multiple sheets
write_xlsx(list(data_crane = data_crane,
                 additional_measurements = data_crane_additional),
           "tmp/data_crane_all.xlsx")

```

Optional exercise (++) - Read and write SPSS, SAS, and STATA data files

```

library(haven) # to read and write SPSS, STATA and SAS files

# create a directory
if (!dir.exists('tmp')){
  dir.create("tmp")
}

# read and write SPSS file
write_sav(data_crane, file.path("tmp", "crane_spss.sav"))
read_sav(file.path("tmp", "crane_spss.sav"))

## # A tibble: 20,873 x 42
##   event_id visible timestamp      location_long location_lat
##   <dbl>    <dbl> <dttm>          <dbl>        <dbl>
## 1 3.79e9     1 2017-10-15 00:00:16      40.3       55.3
## 2 3.80e9     1 2017-10-15 00:00:16      40.3       55.3
## 3 3.79e9     1 2017-10-15 02:00:23      40.3       55.3
## 4 3.80e9     1 2017-10-15 02:00:23      40.3       55.3
## 5 3.79e9     1 2017-10-15 04:00:23      40.3       55.3
## 6 3.80e9     1 2017-10-15 04:00:23      40.3       55.3
## 7 3.79e9     1 2017-10-15 05:00:14      40.1       55.3
## 8 3.80e9     1 2017-10-15 05:00:14      40.1       55.3
## 9 3.80e9     1 2017-10-15 06:00:23      40.1       55.3
## 10 3.80e9    1 2017-10-15 08:00:18      40.1       55.3
## # ... with 20,863 more rows, and 37 more variables:
## #   acceleration_raw_x <dbl>, acceleration_raw_y <dbl>,
## #   acceleration_raw_z <dbl>, bar_barometric_height <dbl>,
## #   battery_charge_percent <dbl>, battery_charging_current <dbl>,
## #   eobs_activity <dbl>, eobs_activity_samples <dbl>,
## #   eobs_battery_voltage <dbl>, eobs_fix_battery_voltage <dbl>,
## #   eobs_horizontal_accuracy_estimate <dbl>, eobs_key_bin_checksum <dbl>,
## #   eobs_speed_accuracy_estimate <dbl>, eobs_start_timestamp <chr>,
## #   eobs_status <chr>, eobs_temperature <dbl>, eobs_type_of_fix <dbl>,
## #   eobs_used_time_to_get_fix <dbl>, external_temperature <dbl>,
## #   gps_hdop <dbl>, gps_satellite_count <dbl>, gps_time_to_fix <dbl>,
## #   ground_speed <dbl>, heading <dbl>, height_above_ellipsoid <dbl>,
## #   height_above_msl <dbl>, import_marked_outlier <dbl>,
## #   gls_light_level <dbl>, mag_magnetic_field_raw_x <dbl>,
## #   mag_magnetic_field_raw_y <dbl>, mag_magnetic_field_raw_z <dbl>,
## #   orn_transmission_protocol <chr>, tag_voltage <dbl>, sensor_type <chr>,
## #   individual_taxon_canonical_name <chr>,
## #   individual_local_identifier <chr>, study_name <chr>

```

```

# The SAS file will not write because a column name is too long
# So we perform a quick fix to generate shorter but unique column names
data_crane_sas <- data_crane
names(data_crane_sas) <- str_c(
  str_sub(names(data_crane_sas), 1, 10),
  str_sub(names(data_crane_sas), -5),
  sep = "_")

# read and write SAS file
write_sas(data_crane_sas, file.path("tmp", "crane_sas.sas7bdat"))
read_sas(file.path("tmp", "crane_sas.sas7bdat"))

## # A tibble: 20,873 x 42
##   event_id_nt_id visible_sible timestamp_stamp      location_l__long
##   <dbl>          <dbl>    <dttm>                  <dbl>
## 1 3794510958        1 2017-10-15 00:00:16      40.3
## 2 3796034354        1 2017-10-15 00:00:16      40.3
## 3 3794510962        1 2017-10-15 02:00:23      40.3
## 4 3796034355        1 2017-10-15 02:00:23      40.3
## 5 3794510961        1 2017-10-15 04:00:23      40.3
## 6 3796034356        1 2017-10-15 04:00:23      40.3
## 7 3794510960        1 2017-10-15 05:00:14      40.1
## 8 3796034357        1 2017-10-15 05:00:14      40.1
## 9 3796034358        1 2017-10-15 06:00:23      40.1
## 10 3796034359       1 2017-10-15 08:00:18      40.1
## # ... with 20,863 more rows, and 38 more variables:
## #   location_l_n_lat <dbl>, accelerati_raw_x <dbl>,
## #   accelerati_raw_y <dbl>, accelerati_raw_z <dbl>,
## #   bar_barome_eight <dbl>, battery_ch_rcent <dbl>,
## #   battery_ch_rrrent <dbl>, eobs_activ_ivity <dbl>,
## #   eobs_activ_mples <dbl>, eobs_batte_ltage <dbl>,
## #   eobs_fix_b_ltage <dbl>, eobs_horiz_imate <dbl>,
## #   eobs_key_b_cksum <dbl>, eobs_speed_imate <dbl>,
## #   eobs_start_stamp <chr>, eobs_statu_status <chr>,
## #   eobs_tempe_ature <dbl>, eobs_type_f_fix <dbl>,
## #   eobs_used_t_fix <dbl>, external_t_ature <dbl>, gps_hdop_hdop <dbl>,
## #   gps_satell_count <dbl>, gps_time_t_o_fix <dbl>,
## #   ground_spe_speed <dbl>, heading_ading <dbl>, height_abo_psoid <dbl>,
## #   height_abo_e_msl <dbl>, import_mar_tlier <dbl>,
## #   gls_light_level <dbl>, mag_magnet_raw_x <dbl>,
## #   mag_magnet_raw_y <dbl>, mag_magnet_raw_z <dbl>,
## #   orn_transm_tocol <chr>, tag_voltag_ltage <dbl>,
## #   sensor_typ_type <chr>, individual_name <chr>,
## #   individual_ifier <chr>, study_name_name <chr>

# read and write STATA file
write_dta(data_crane, file.path("tmp", "crane_stata.dta"))
read_dta(file.path("tmp", "crane_stata.dta"))

## # A tibble: 20,873 x 42
##   event_id visible timestamp      location_long location_lat
##   <dbl>     <dbl>    <dttm>                  <dbl>        <dbl>
## 1 3.79e9      1 2017-10-15 00:00:16      40.3        55.3
## 2 3.80e9      1 2017-10-15 00:00:16      40.3        55.3

```

```

## 3 3.79e9      1 2017-10-15 02:00:23      40.3      55.3
## 4 3.80e9      1 2017-10-15 02:00:23      40.3      55.3
## 5 3.79e9      1 2017-10-15 04:00:23      40.3      55.3
## 6 3.80e9      1 2017-10-15 04:00:23      40.3      55.3
## 7 3.79e9      1 2017-10-15 05:00:14      40.1      55.3
## 8 3.80e9      1 2017-10-15 05:00:14      40.1      55.3
## 9 3.80e9      1 2017-10-15 06:00:23      40.1      55.3
## 10 3.80e9     1 2017-10-15 08:00:18      40.1      55.3
## # ... with 20,863 more rows, and 37 more variables:
## #   acceleration_raw_x <dbl>, acceleration_raw_y <dbl>,
## #   acceleration_raw_z <dbl>, bar_barometric_height <dbl>,
## #   battery_charge_percent <dbl>, battery_charging_current <dbl>,
## #   eobs_activity <dbl>, eobs_activity_samples <dbl>,
## #   eobs_battery_voltage <dbl>, eobs_fix_battery_voltage <dbl>,
## #   eobs_horizontal_accuracy_estimate <dbl>, eobs_key_bin_checksum <dbl>,
## #   eobs_speed_accuracy_estimate <dbl>, eobs_start_timestamp <chr>,
## #   eobs_status <chr>, eobs_temperature <dbl>, eobs_type_of_fix <dbl>,
## #   eobs_used_time_to_get_fix <dbl>, external_temperature <dbl>,
## #   gps_hdop <dbl>, gps_satellite_count <dbl>, gps_time_to_fix <dbl>,
## #   ground_speed <dbl>, heading <dbl>, height_above_ellipsoid <dbl>,
## #   height_above_msl <dbl>, import_marked_outlier <dbl>,
## #   gls_light_level <dbl>, mag_magnetic_field_raw_x <dbl>,
## #   mag_magnetic_field_raw_y <dbl>, mag_magnetic_field_raw_z <dbl>,
## #   orn_transmission_protocol <chr>, tag_voltage <dbl>, sensor_type <chr>,
## #   individual_taxon_canonical_name <chr>,
## #   individual_local_identifier <chr>, study_name <chr>

```

Optional exercise (++) - Parse datetime columns

```

glimpse(data_crane)

## Observations: 20,873
## Variables: 42
## $ event_id                  <dbl> 3794510958, 3796034354, 3794...
## $ visible                   <lgl> TRUE, TRUE, TRUE, TRUE, TRUE...
## $ timestamp                 <dttm> 2017-10-15 00:00:16, 2017-1...
## $ location_long              <dbl> 40.31509, 40.31509, 40.31481...
## $ location_lat               <dbl> 55.34764, 55.34764, 55.34792...
## $ acceleration_raw_x         <dbl> NA, NA, NA, NA, NA, NA, ...
## $ acceleration_raw_y         <dbl> NA, NA, NA, NA, NA, NA, ...
## $ acceleration_raw_z         <dbl> NA, NA, NA, NA, NA, NA, ...
## $ bar_barometric_height      <dbl> NA, NA, NA, NA, NA, NA, ...
## $ battery_charge_percent     <dbl> NA, NA, NA, NA, NA, NA, ...
## $ battery_charging_current   <dbl> NA, NA, NA, NA, NA, NA, ...
## $ eobs_activity               <lgl> NA, NA, NA, NA, NA, NA, ...
## $ eobs_activity_samples       <lgl> NA, NA, NA, NA, NA, NA, ...
## $ eobs_battery_voltage        <dbl> 3784, 3818, 3784, 3818, 3784...
## $ eobs_fix_battery_voltage    <dbl> NA, 3801, NA, 3798, NA, 3796...
## $ eobs_horizontal_accuracy_estimate <dbl> NA, 48.90, NA, 2.56, NA, 3.8...
## $ eobs_key_bin_checksum       <dbl> 0, 2787211109, 0, 1067160123...
## $ eobs_speed_accuracy_estimate <dbl> NA, 1.54, NA, 0.33, NA, 0.34...
## $ eobs_start_timestamp        <chr> NA, "2017,15+Oct", NA, "2017...
## $ eobs_status                <chr> NA, "A", NA, "A", NA, "A", N...

```

```

## $ eobs_temperature <dbl> NA, -7, NA, -7, NA, -8, NA, ...
## $ eobs_type_of_fix <dbl> NA, 3, NA, 3, NA, 3, NA, 3, ...
## $ eobs_used_time_to_get_fix <dbl> NA, 15, NA, 22, NA, 21, NA, ...
## $ external_temperature <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ gps_hdop <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ gps_satellite_count <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ gps_time_to_fix <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ ground_speed <dbl> NA, 0.34, NA, 0.01, NA, 12.8...
## $ heading <dbl> NA, 340.13, NA, 0.00, NA, 22...
## $ height_above_ellipsoid <dbl> NA, 127.0, NA, 112.2, NA, 20...
## $ height_above_msl <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ import_marked_outlier <lgl> NA, NA, NA, NA, NA, NA, NA, ...
## $ gls_light_level <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ mag_magnetic_field_raw_x <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ mag_magnetic_field_raw_y <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ mag_magnetic_field_raw_z <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ orn_transmission_protocol <chr> NA, NA, NA, NA, NA, NA, NA, ...
## $ tag_voltage <dbl> NA, NA, NA, NA, NA, NA, NA, ...
## $ sensor_type <chr> "gps", "gps", "gps", "gps", ...
## $ individual_taxon_canonical_name <chr> "Grus grus", "Grus grus", "G...
## $ individual_local_identifier <chr> "L6037", "L6037", "L6037", ...
## $ study_name <chr> "Huji JNF Crane israel GPRS"...

```

```

read_csv('data/HUJI_Crane_Israel_GPRS.csv',
  col_types = cols(
    . = col_guess(),
    eobs_start_timestamp = col_date(format="%Y,%d+%b")
  ))

```

Warning: The following named parsers don't match the column names: .

```

## # A tibble: 20,873 x 42
##   event_id visible timestamp      location_long location_lat
##   <dbl> <lgl>   <dttm>          <dbl>           <dbl>
## 1 3.79e9  TRUE   2017-10-15 00:00:16     40.3        55.3
## 2 3.80e9  TRUE   2017-10-15 00:00:16     40.3        55.3
## 3 3.79e9  TRUE   2017-10-15 02:00:23     40.3        55.3
## 4 3.80e9  TRUE   2017-10-15 02:00:23     40.3        55.3
## 5 3.79e9  TRUE   2017-10-15 04:00:23     40.3        55.3
## 6 3.80e9  TRUE   2017-10-15 04:00:23     40.3        55.3
## 7 3.79e9  TRUE   2017-10-15 05:00:14     40.1        55.3
## 8 3.80e9  TRUE   2017-10-15 05:00:14     40.1        55.3
## 9 3.80e9  TRUE   2017-10-15 06:00:23     40.1        55.3
## 10 3.80e9 TRUE   2017-10-15 08:00:18     40.1        55.3
## # ... with 20,863 more rows, and 37 more variables:
## #   acceleration_raw_x <dbl>, acceleration_raw_y <dbl>,
## #   acceleration_raw_z <dbl>, bar_barometric_height <dbl>,
## #   battery_charge_percent <dbl>, battery_charging_current <dbl>,
## #   eobs_activity <lgl>, eobs_activity_samples <lgl>,
## #   eobs_battery_voltage <dbl>, eobs_fix_battery_voltage <dbl>,
## #   eobs_horizontal_accuracy_estimate <dbl>, eobs_key_bin_checksum <dbl>,
## #   eobs_speed_accuracy_estimate <dbl>, eobs_start_timestamp <date>,
## #   eobs_status <chr>, eobs_temperature <dbl>, eobs_type_of_fix <dbl>,
## #   eobs_used_time_to_get_fix <dbl>, external_temperature <dbl>,
## #   gps_hdop <dbl>, gps_satellite_count <dbl>, gps_time_to_fix <dbl>,

```

```

## #  ground_speed <dbl>, heading <dbl>, height_above_ellipsoid <dbl>,
## #  height_above_msl <dbl>, import_marked_outlier <lgl>,
## #  gls_light_level <dbl>, mag_magnetic_field_raw_x <dbl>,
## #  mag_magnetic_field_raw_y <dbl>, mag_magnetic_field_raw_z <dbl>,
## #  orn_transmission_protocol <chr>, tag_voltage <dbl>, sensor_type <chr>,
## #  individual_taxon_canonical_name <chr>,
## #  individual_local_identifier <chr>, study_name <chr>

```

2. Data visualisation

Basic exercise I - Quick plots of the `data_crane`.

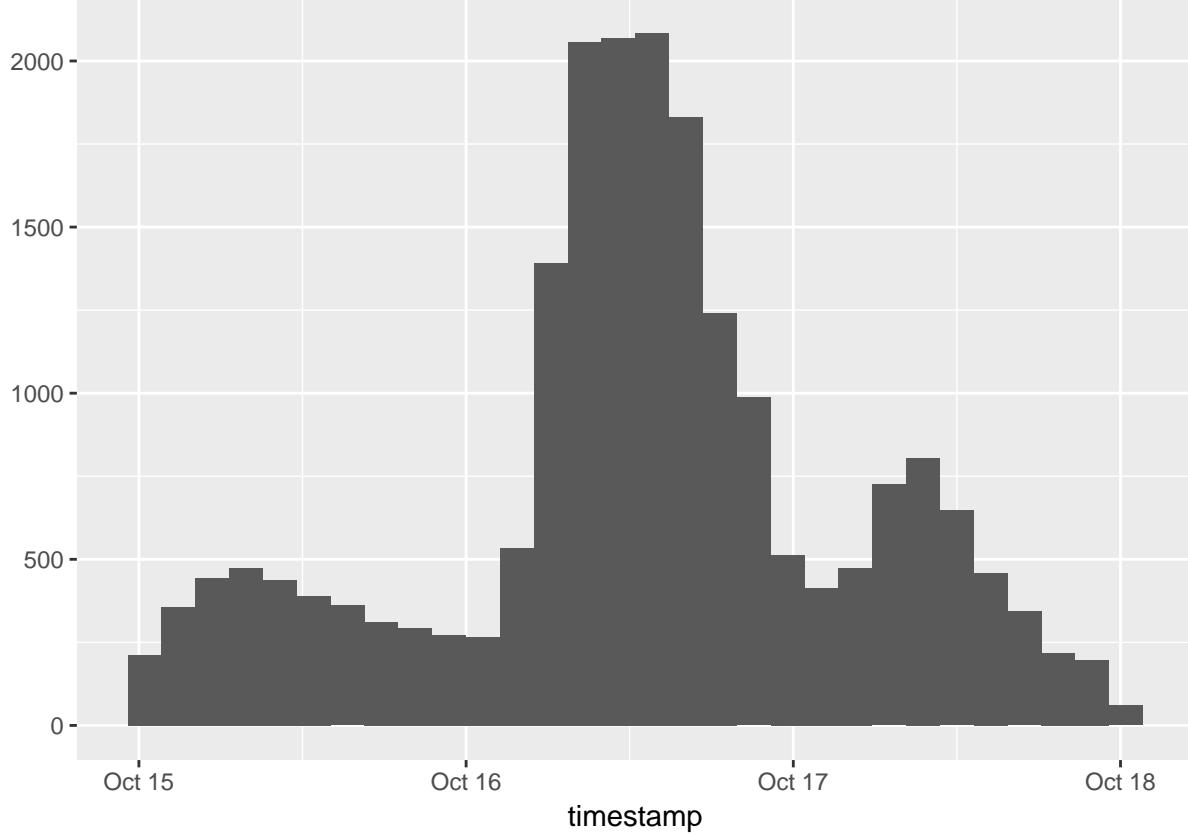
a) Single column plots

```

# for example...
qplot(timestamp, data=data_crane)

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

```



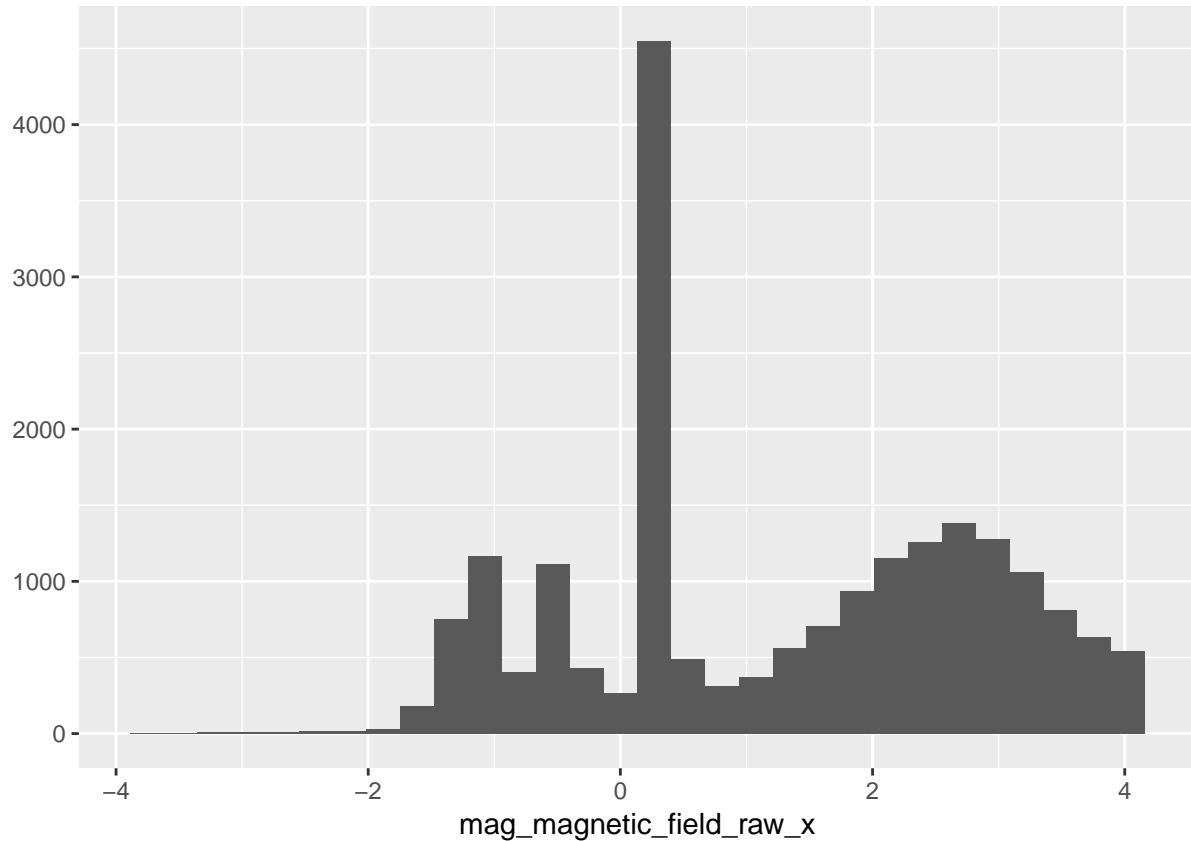
```

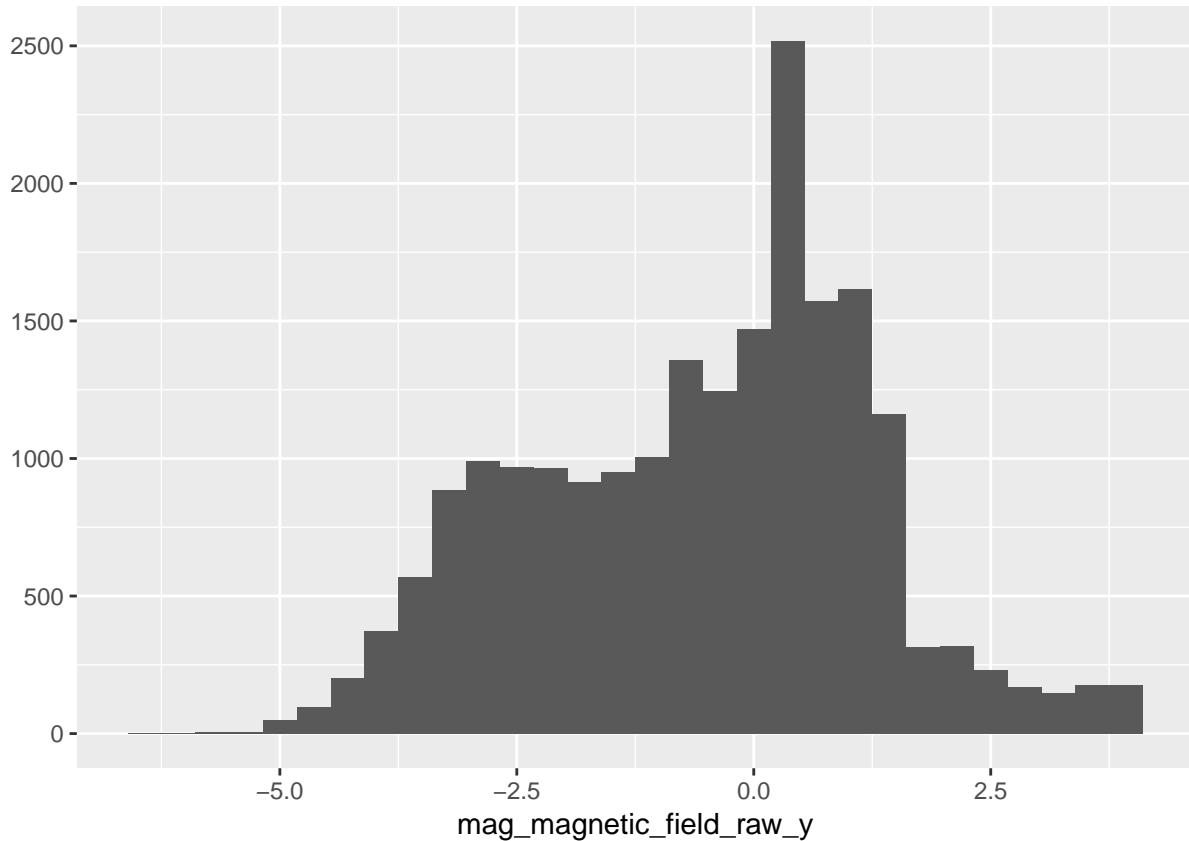
qplot(mag_magnetic_field_raw_x, data=data_crane)

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

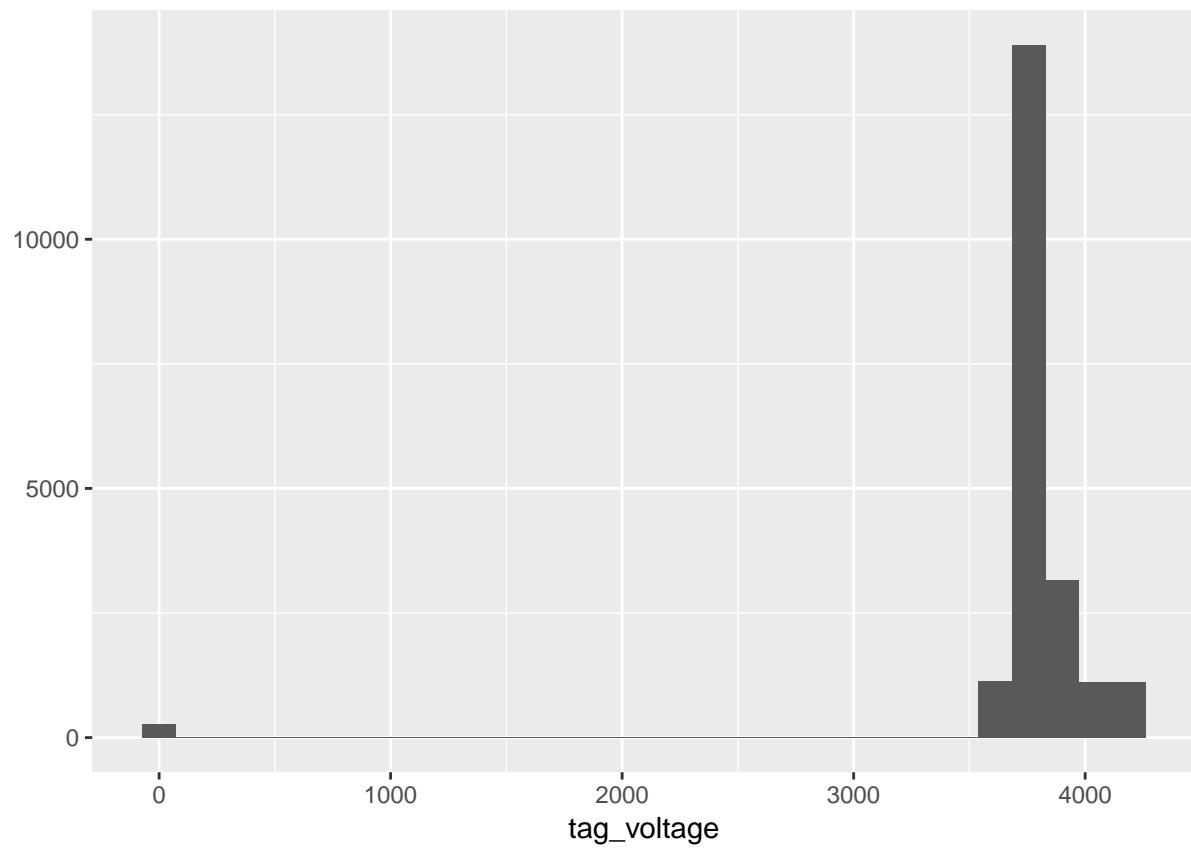
## Warning: Removed 432 rows containing non-finite values (stat_bin).

```

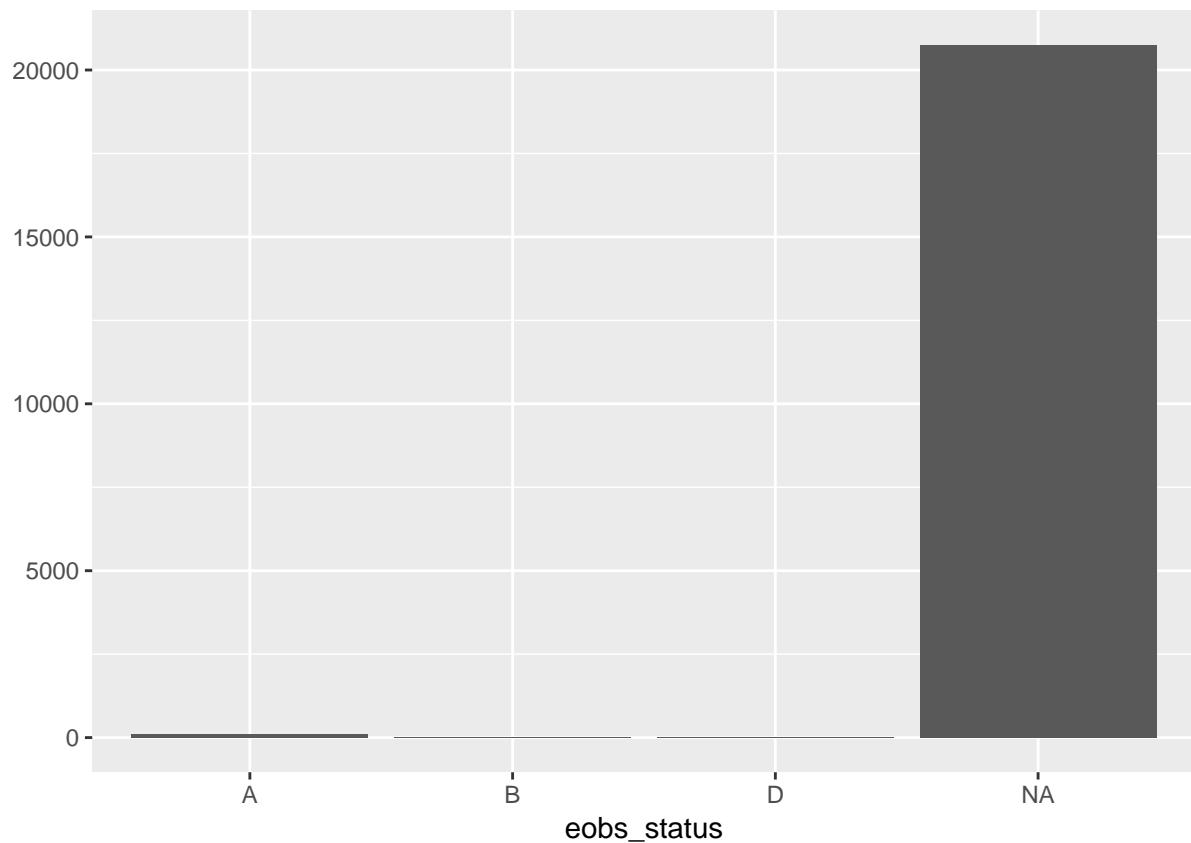




```
qplot(tag_voltage, data=data_crane)  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## Warning: Removed 155 rows containing non-finite values (stat_bin).
```



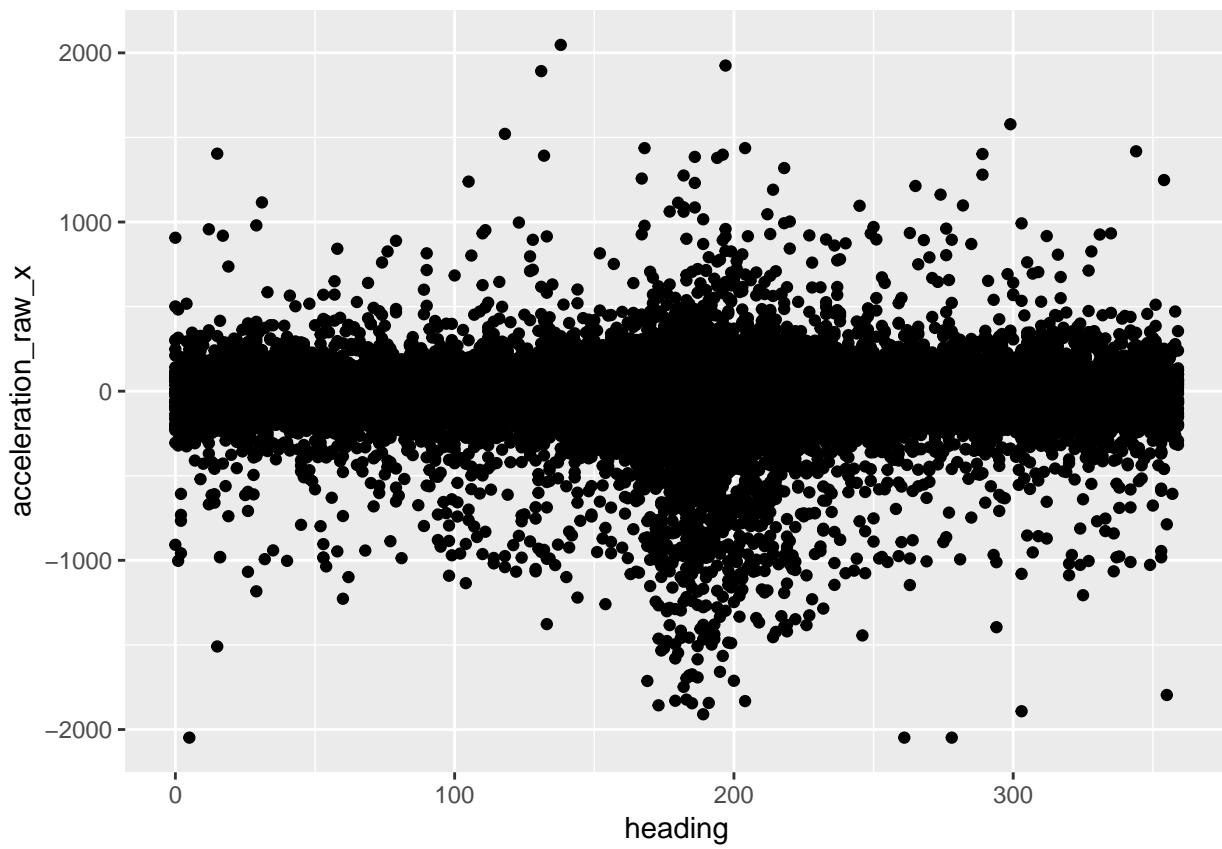
```
qplot(eobs_status, data=data_crane)
```



b) Two column plots

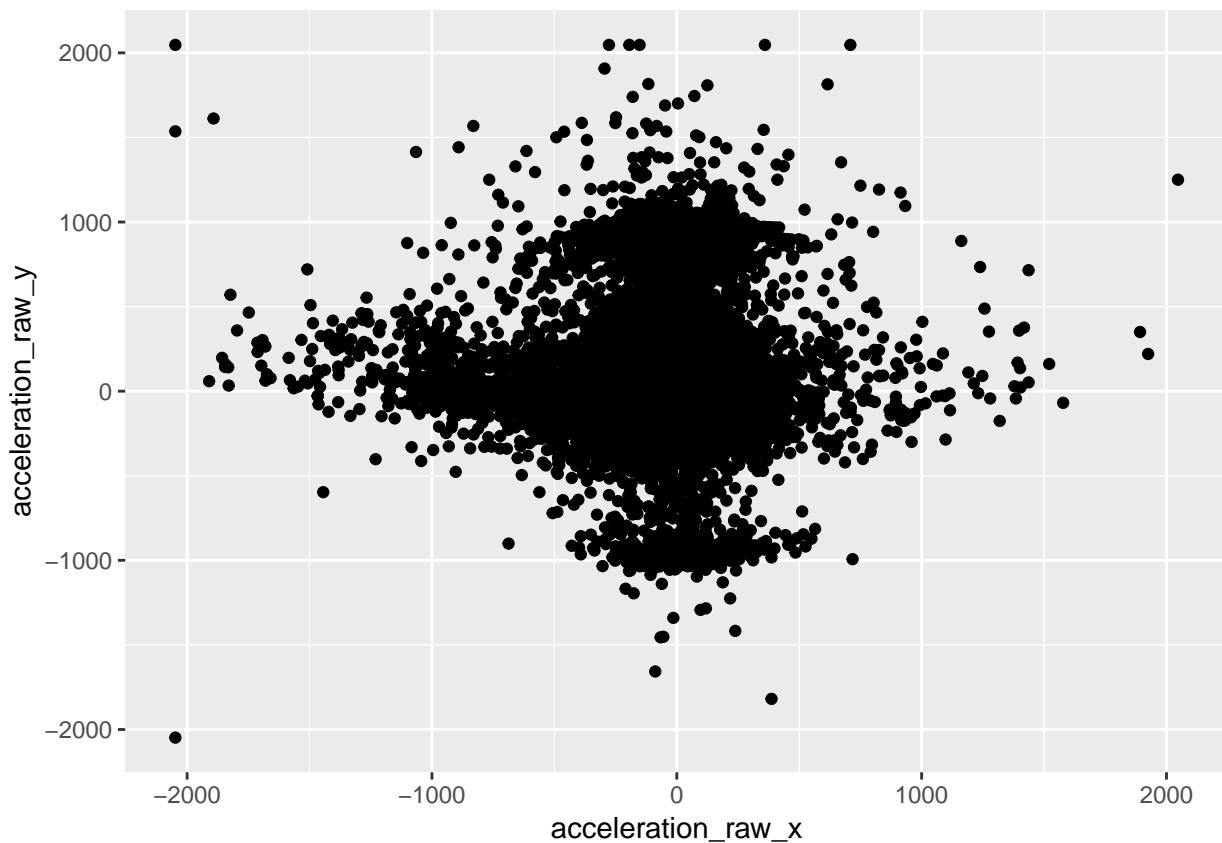
```
# for example...
qplot(heading, acceleration_raw_x, data=data_crane)

## Warning: Removed 432 rows containing missing values (geom_point).
```

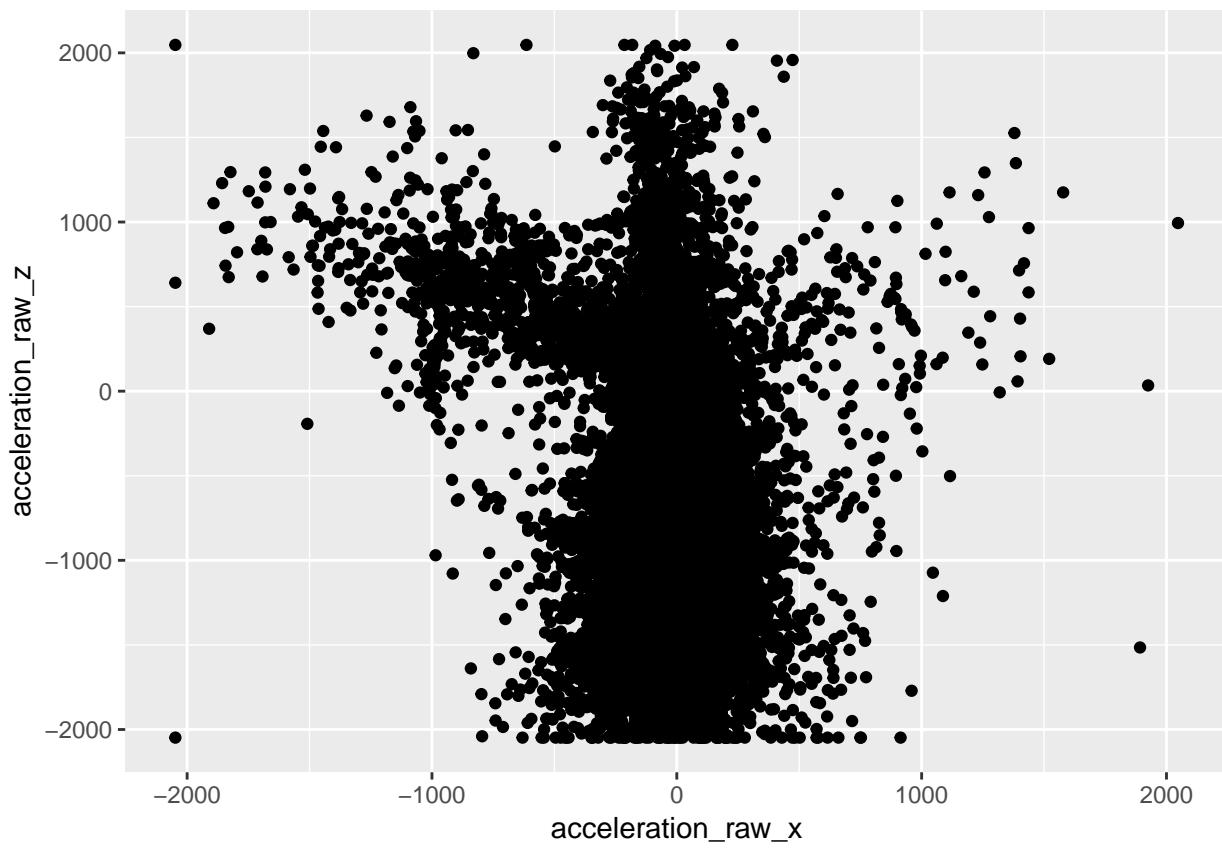


```
qplot(acceleration_raw_x, acceleration_raw_y, data=data_crane)
```

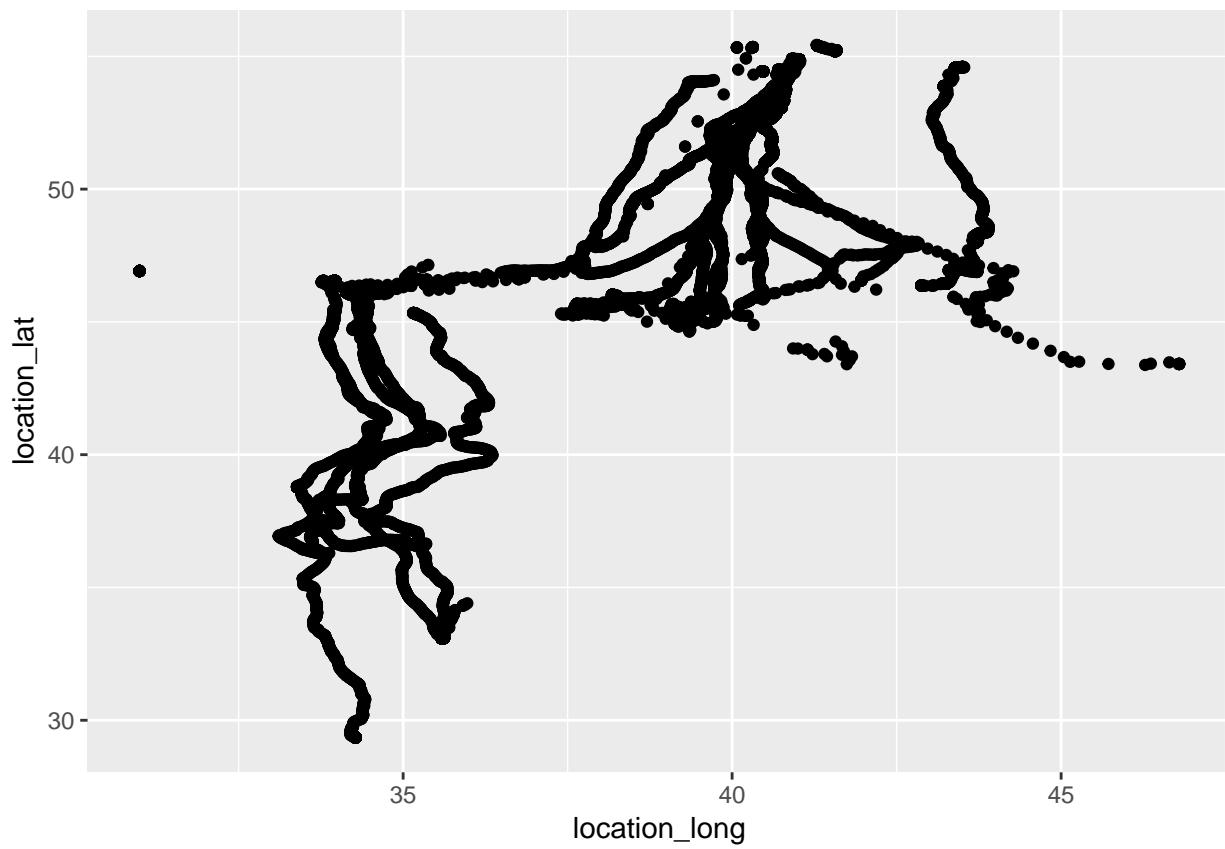
```
## Warning: Removed 432 rows containing missing values (geom_point).
```



```
## Warning: Removed 432 rows containing missing values (geom_point).
```

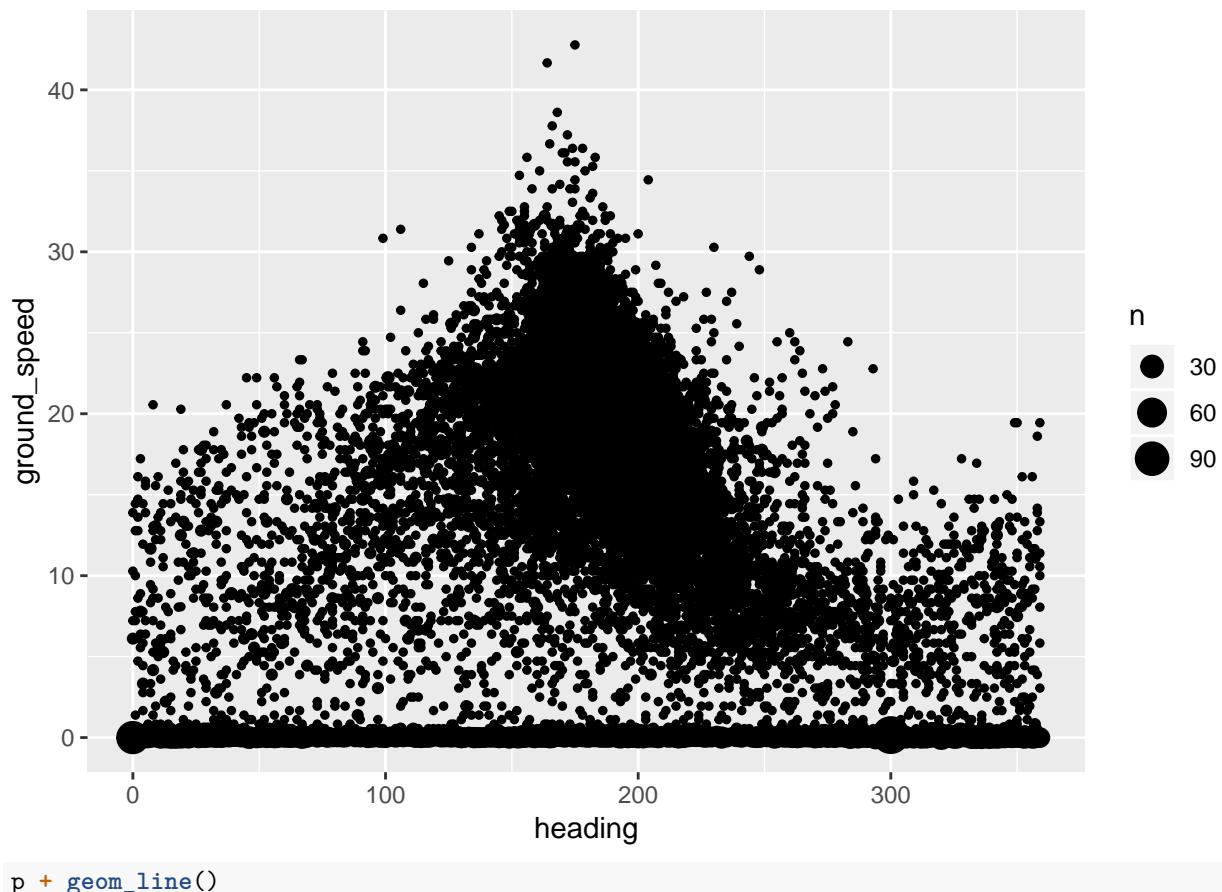


```
## Warning: Removed 200 rows containing missing values (geom_point).
```



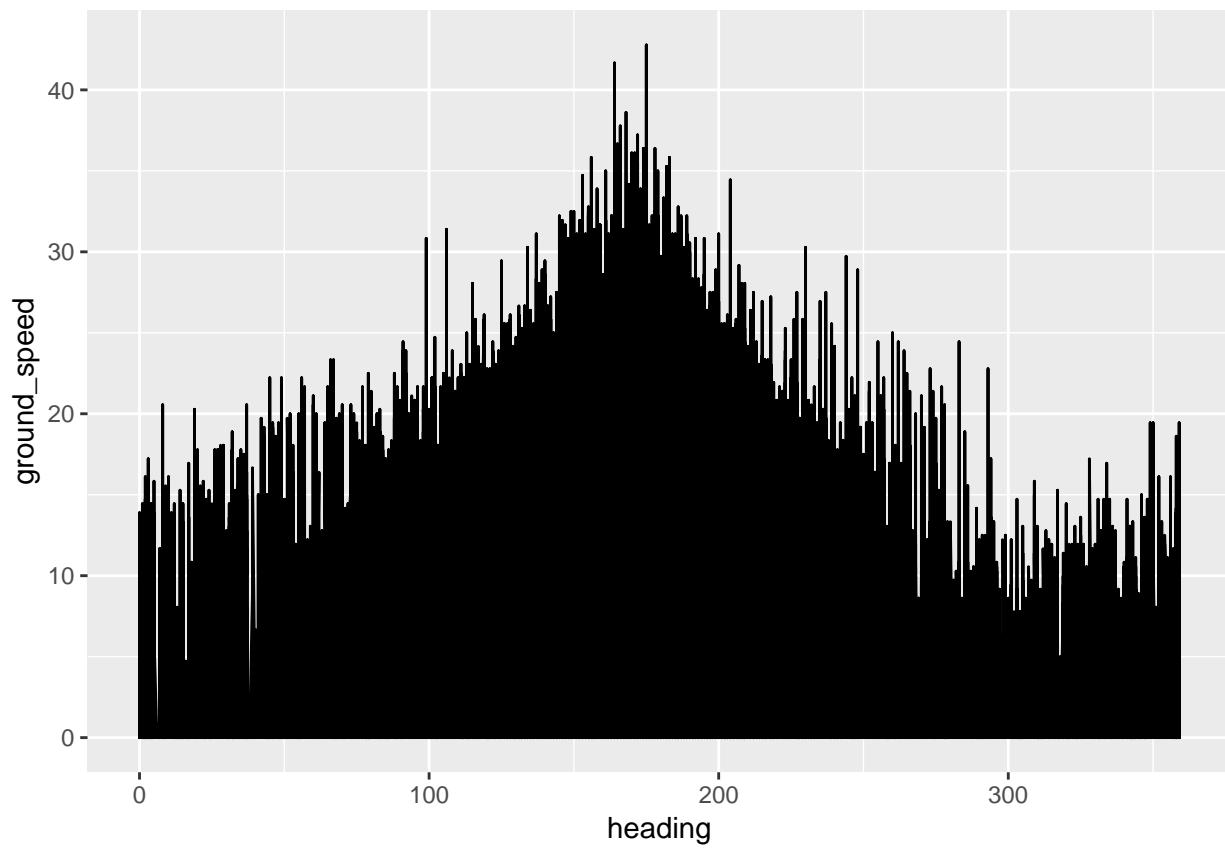
Basic exercise II - Using ggplot for plotting

```
p <- ggplot(data_crane, aes(x = heading, y = ground_speed))  
p + geom_count()  
  
## Warning: Removed 40 rows containing non-finite values (stat_sum).
```



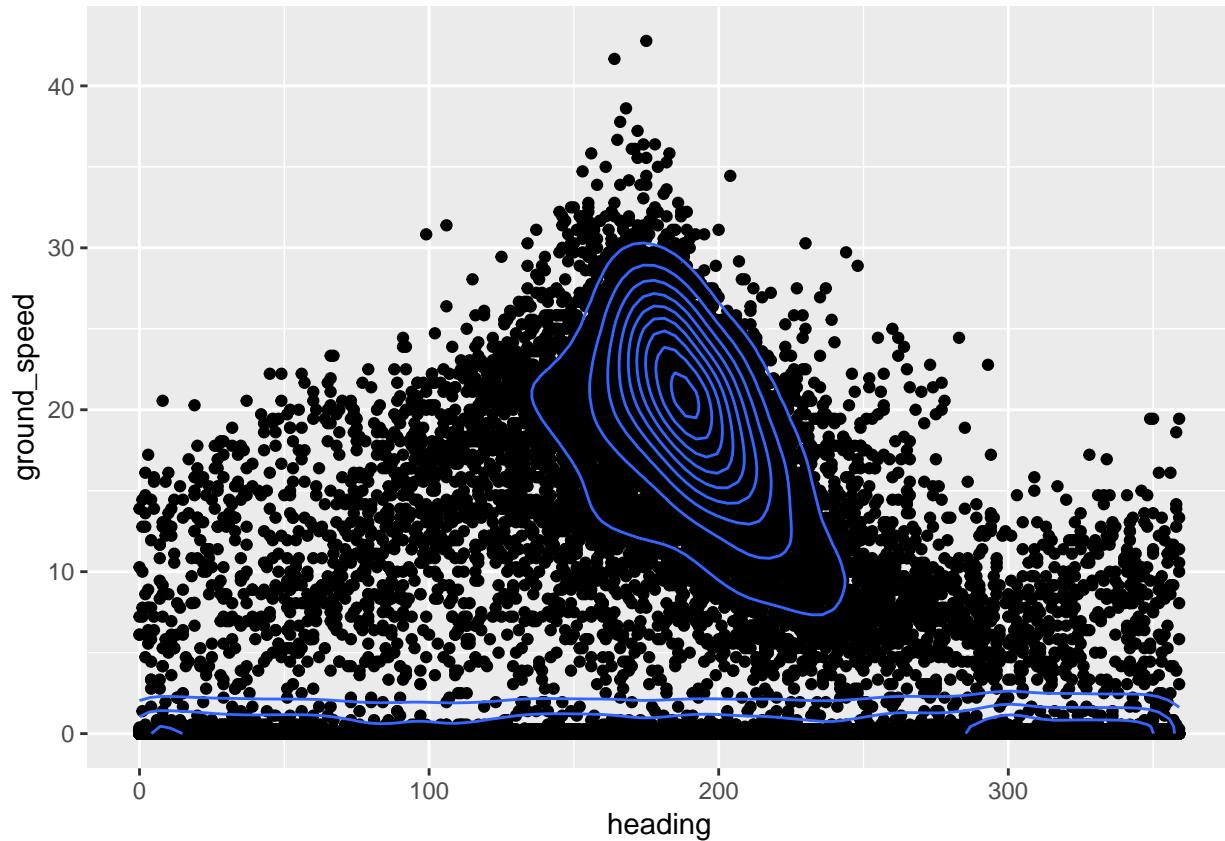
```
p + geom_line()
```

```
## Warning: Removed 40 rows containing missing values (geom_path).
```



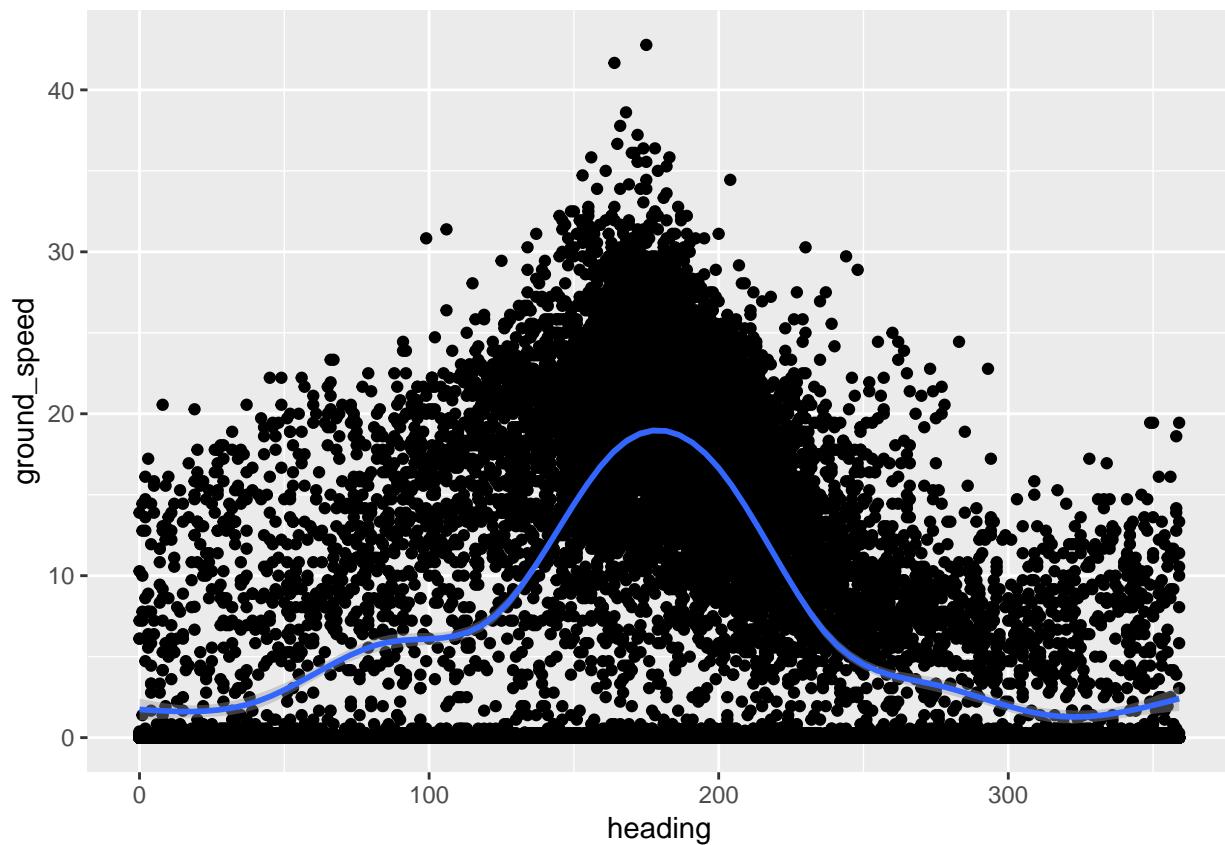
```
p + geom_point() + geom_density_2d()
```

```
## Warning: Removed 40 rows containing non-finite values (stat_density2d).  
## Warning: Removed 40 rows containing missing values (geom_point).
```



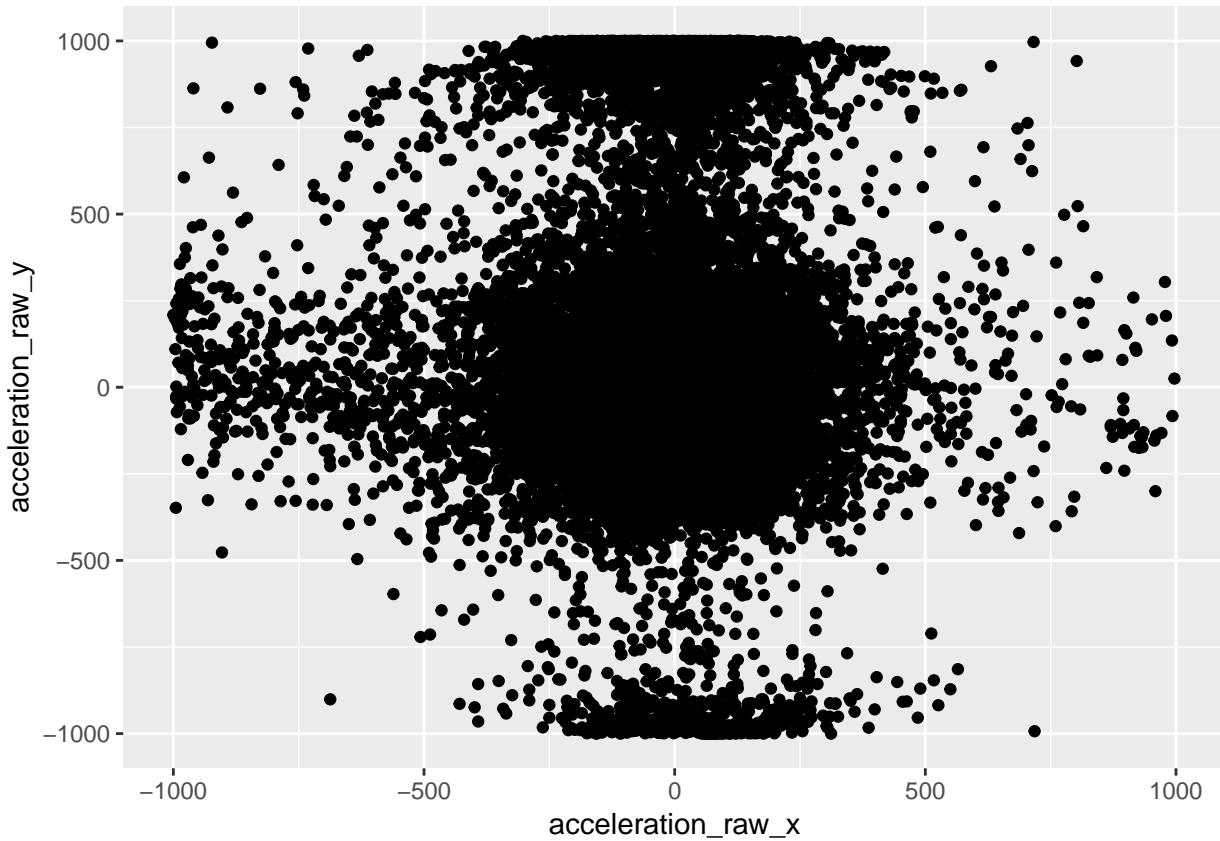
Optional exercise (+) - Statistical layers for graphs.

```
p + geom_point() + geom_smooth()  
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'  
## Warning: Removed 40 rows containing non-finite values (stat_smooth).  
## Warning: Removed 40 rows containing missing values (geom_point).
```



Optional exercise (+) - Scale axes

```
ggplot(data_crane, aes(acceleration_raw_x, acceleration_raw_y)) +  
  geom_point() +  
  scale_x_continuous(limits = c(-1000, 1000)) +  
  scale_y_continuous(limits = c(-1000, 1000))  
  
## Warning: Removed 2436 rows containing missing values (geom_point).
```



Optional exercise (++) - Plot the crane positions on a map

a) Install the package maps

```
install.packages('maps')

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
library(maps)

##
## Attaching package: 'maps'
## The following object is masked from 'package:purrr':
## 
##     map
```

b) Plot the crane data on a map.

```
world_map_polygon <- map_data("world2")

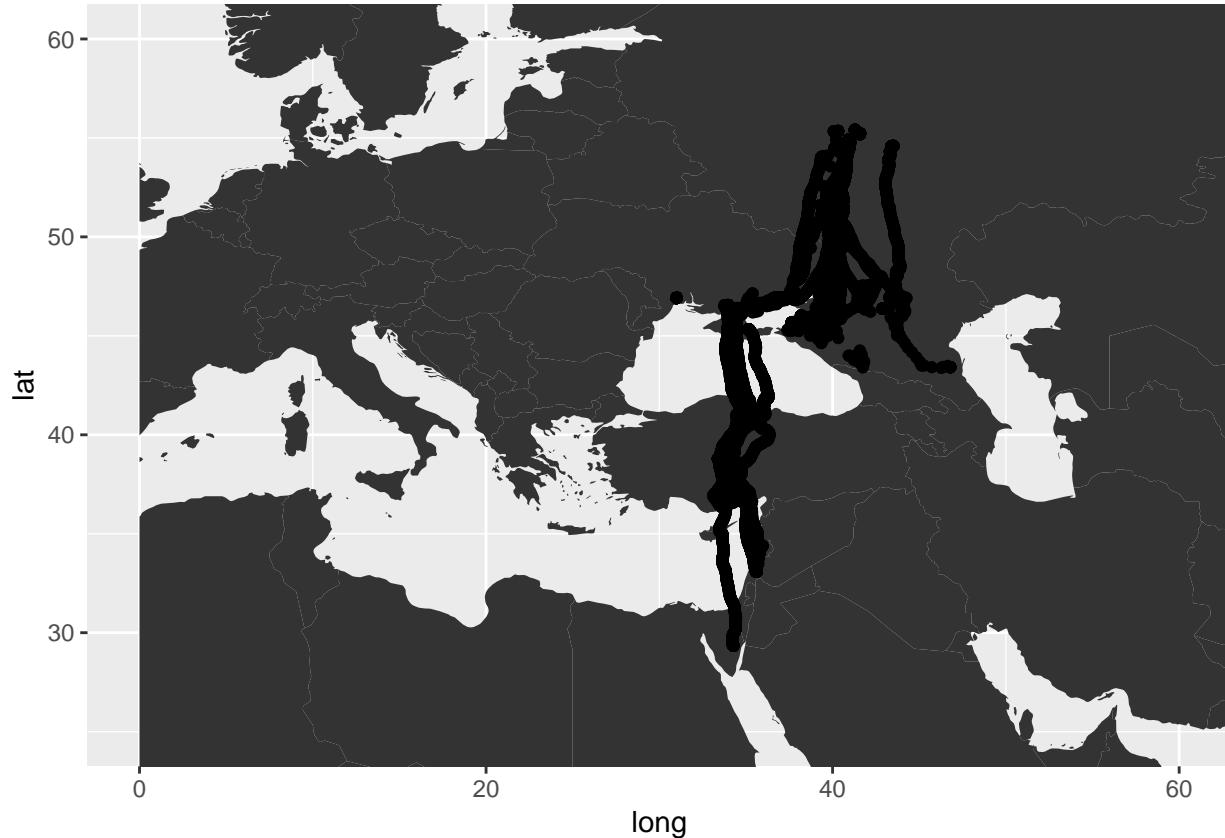
ggplot(data_crane) +
  geom_map(data = world_map_polygon,
           map= world_map_polygon,
           aes(x=long, y = lat, map_id = region)) +
```

```

scale_x_continuous(limits = c(0, 60)) +
scale_y_continuous(limits = c(25, 60)) +
geom_point(data = data_crane,
aes(x = location_long, y = location_lat))

## Warning: Ignoring unknown aesthetics: x, y
## Warning: Removed 200 rows containing missing values (geom_point).

```



c) Use an individual identifier to colour the different cranes.

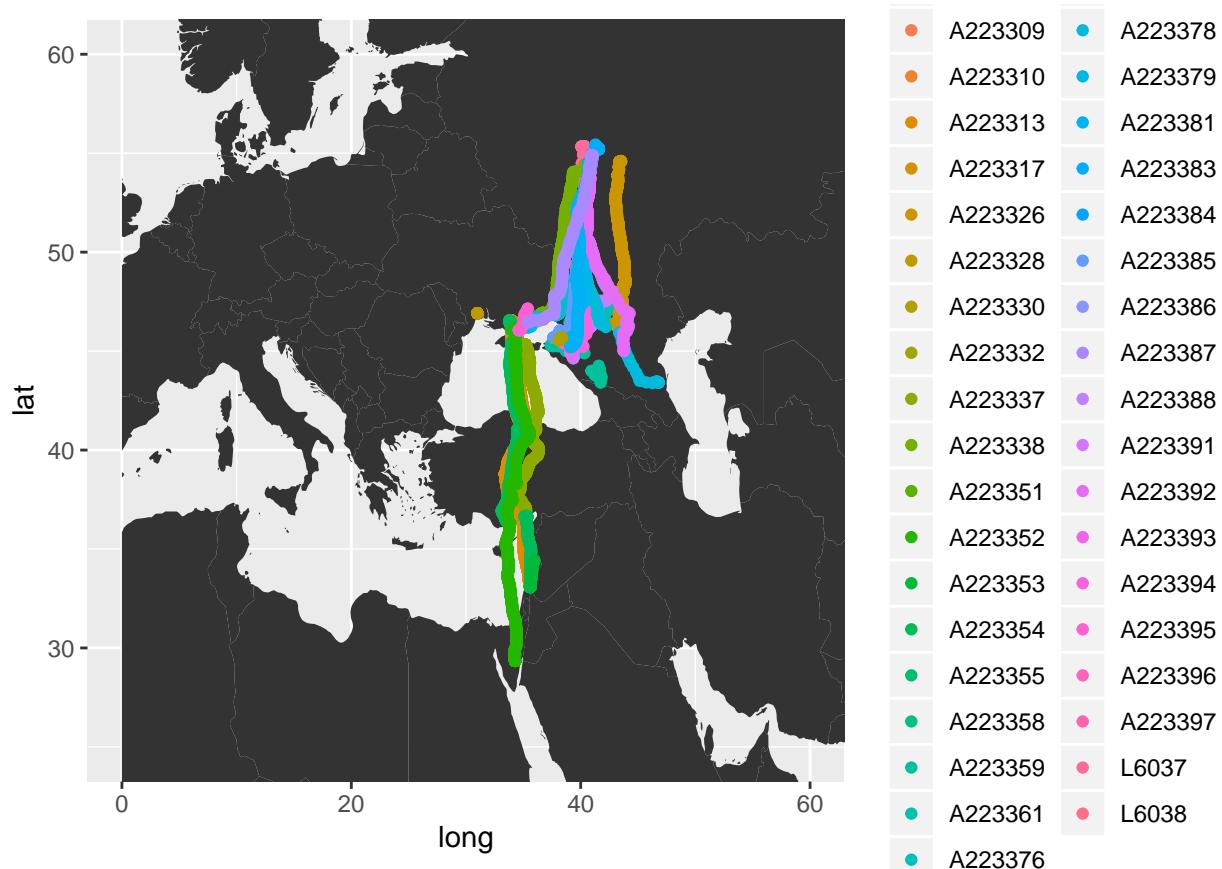
```

world_map_polygon <- map_data("world2")

ggplot(data_crane) +
  geom_map(data = world_map_polygon,
    map= world_map_polygon,
    aes(long, lat, map_id = region)) +
  scale_x_continuous(limits = c(0, 60)) +
  scale_y_continuous(limits = c(25, 60)) +
  geom_point(data = data_crane,
    aes(location_long,
    location_lat,
    colour=individual_local_identifier))

## Warning: Ignoring unknown aesthetics: x, y
## Warning: Removed 200 rows containing missing values (geom_point).

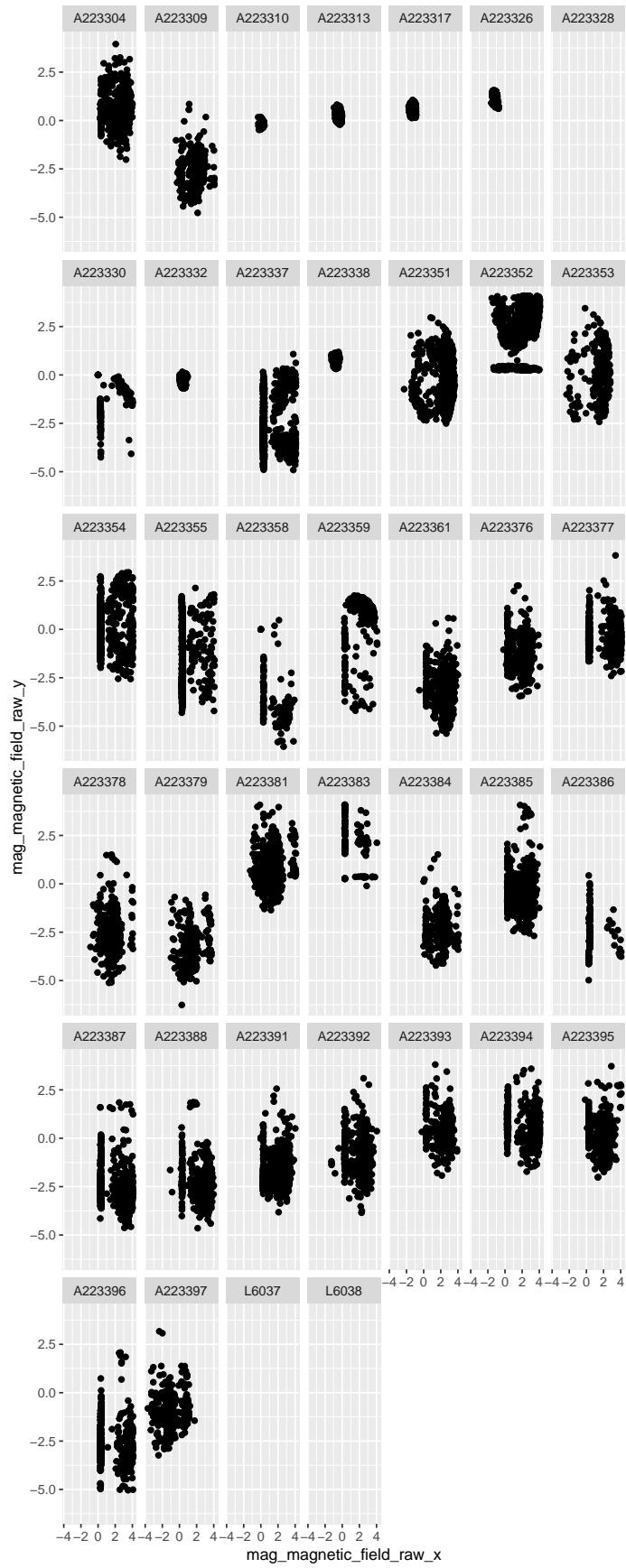
```



Optional exercise (++) - Create facets.

```
ggplot(data_crane, aes(mag_magnetic_field_raw_x, mag_magnetic_field_raw_y)) +
  geom_point() +
  facet_wrap(~individual_local_identifier)

## Warning: Removed 432 rows containing missing values (geom_point).
```



3. Data transformation

Basic exercise I - Subset data

a) Filter data from one individual crane

```
filter(data_crane, individual_local_identifier=='L6037')

## # A tibble: 61 x 42
##   event_id visible timestamp      location_long location_lat
##   <dbl>    <lgl>   <dttm>          <dbl>        <dbl>
## 1 3.79e9  TRUE  2017-10-15 00:00:16     40.3       55.3
## 2 3.80e9  TRUE  2017-10-15 00:00:16     40.3       55.3
## 3 3.79e9  TRUE  2017-10-15 02:00:23     40.3       55.3
## 4 3.80e9  TRUE  2017-10-15 02:00:23     40.3       55.3
## 5 3.79e9  TRUE  2017-10-15 04:00:23     40.3       55.3
## 6 3.80e9  TRUE  2017-10-15 04:00:23     40.3       55.3
## 7 3.79e9  TRUE  2017-10-15 05:00:14     40.1       55.3
## 8 3.80e9  TRUE  2017-10-15 05:00:14     40.1       55.3
## 9 3.80e9  TRUE  2017-10-15 06:00:23     40.1       55.3
## 10 3.80e9 TRUE  2017-10-15 08:00:18     40.1       55.3
## # ... with 51 more rows, and 37 more variables: acceleration_raw_x <dbl>,
## #   acceleration_raw_y <dbl>, acceleration_raw_z <dbl>,
## #   bar_barometric_height <dbl>, battery_charge_percent <dbl>,
## #   battery_charging_current <dbl>, eobs_activity <lgl>,
## #   eobs_activity_samples <lgl>, eobs_battery_voltage <dbl>,
## #   eobs_fix_battery_voltage <dbl>,
## #   eobs_horizontal_accuracy_estimate <dbl>, eobs_key_bin_checksum <dbl>,
## #   eobs_speed_accuracy_estimate <dbl>, eobs_start_timestamp <chr>,
## #   eobs_status <chr>, eobs_temperature <dbl>, eobs_type_of_fix <dbl>,
## #   eobs_used_time_to_get_fix <dbl>, external_temperature <dbl>,
## #   gps_hdop <dbl>, gps_satellite_count <dbl>, gps_time_to_fix <dbl>,
## #   ground_speed <dbl>, heading <dbl>, height_above_ellipsoid <dbl>,
## #   height_above_msl <dbl>, import_marked_outlier <lgl>,
## #   gls_light_level <dbl>, mag_magnetic_field_raw_x <dbl>,
## #   mag_magnetic_field_raw_y <dbl>, mag_magnetic_field_raw_z <dbl>,
## #   orn_transmission_protocol <chr>, tag_voltage <dbl>, sensor_type <chr>,
## #   individual_taxon_canonical_name <chr>,
## #   individual_local_identifier <chr>, study_name <chr>
```

b) Filter data with complete GPS information

```
data_crane_filtered <- filter(data_crane,
                                individual_local_identifier=='L6037',
                                !is.na(eobs_status))

# count the number of rows
nrow(data_crane_filtered)

## [1] 42
```

c) Select specific columns from your data

```
select(data_crane, starts_with("eobs_"))

## # A tibble: 20,873 x 12
##   eobs_activity eobs_activity_s~ eobs_battery_vo~ eobs_fix_batter~
##   <lgl>          <lgl>          <dbl>          <dbl>
## 1 NA            NA             3784           NA
## 2 NA            NA             3818           3801
## 3 NA            NA             3784           NA
## 4 NA            NA             3818           3798
## 5 NA            NA             3784           NA
## 6 NA            NA             3814           3796
## 7 NA            NA             3784           NA
## 8 NA            NA             3814           3798
## 9 NA            NA             3814           3796
## 10 NA           NA             3814           3798
## # ... with 20,863 more rows, and 8 more variables:
## #   eobs_horizontal_accuracy_estimate <dbl>, eobs_key_bin_checksum <dbl>,
## #   eobs_speed_accuracy_estimate <dbl>, eobs_start_timestamp <chr>,
## #   eobs_status <chr>, eobs_temperature <dbl>, eobs_type_of_fix <dbl>,
## #   eobs_used_time_to_get_fix <dbl>
```

d) Combine filter and select using the pipe operator

```
filter(data_crane, individual_local_identifier=='L6037', !is.na(eobs_status)) %>%
  select(starts_with("eobs_"))

## # A tibble: 42 x 12
##   eobs_activity eobs_activity_s~ eobs_battery_vo~ eobs_fix_batter~
##   <lgl>          <lgl>          <dbl>          <dbl>
## 1 NA            NA             3818           3801
## 2 NA            NA             3818           3798
## 3 NA            NA             3814           3796
## 4 NA            NA             3814           3798
## 5 NA            NA             3814           3796
## 6 NA            NA             3814           3798
## 7 NA            NA             3815           3798
## 8 NA            NA             3815           3801
## 9 NA            NA             3811           3792
## 10 NA           NA             3811           3791
## # ... with 32 more rows, and 8 more variables:
## #   eobs_horizontal_accuracy_estimate <dbl>, eobs_key_bin_checksum <dbl>,
## #   eobs_speed_accuracy_estimate <dbl>, eobs_start_timestamp <chr>,
## #   eobs_status <chr>, eobs_temperature <dbl>, eobs_type_of_fix <dbl>,
## #   eobs_used_time_to_get_fix <dbl>
```

Basic exercise II - Compute the magnitude of the magnetic field

```
data_crane_magnetic <- mutate(data_crane,
                                magnetic_magnitude = sqrt(mag_magnetic_field_raw_x ^ 2 +
```

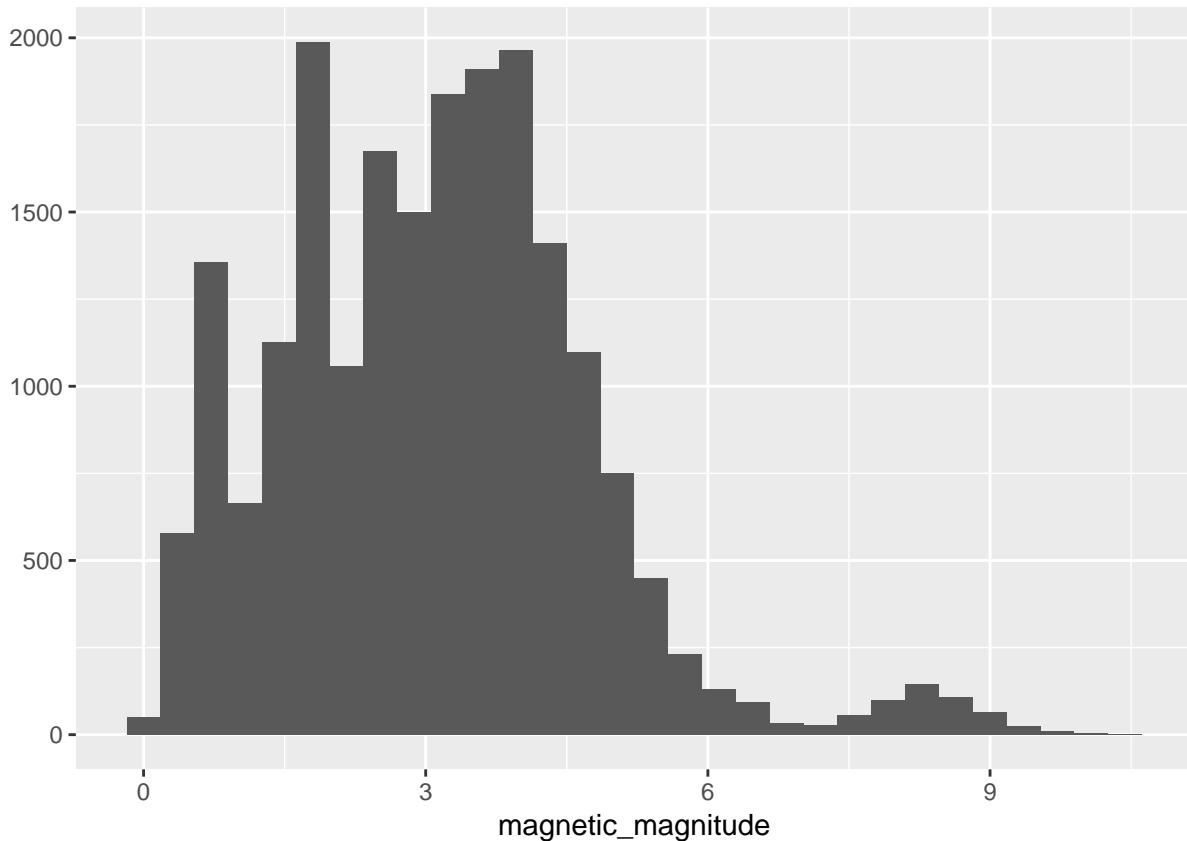
```

mag_magnetic_field_raw_y ^ 2 +
mag_magnetic_field_raw_z ^ 2))

# take a look at the new variable
qplot(magnetic_magnitude, data=data_crane_magnetic)

```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
Warning: Removed 432 rows containing non-finite values (stat_bin).



Optional exercise (+) - Summarise results

```

summarise(data_crane,
  min_latitude = min(location_lat, na.rm=T),
  max_latitude = max(location_lat, na.rm=T),
  first_observation = min(timestamp),
  last_observation = max(timestamp),
  magnitude_acceleration = mean(sqrt(acceleration_raw_x ^ 2 +
                                         acceleration_raw_y ^ 2 +
                                         acceleration_raw_z ^ 2), na.rm = T))

## # A tibble: 1 x 5
##   min_latitude max_latitude first_observation   last_observation
##       <dbl>          <dbl>        <dttm>           <dttm>
## 1      29.3        55.4 2017-10-15 00:00:05 2017-10-17 23:59:33
## # ... with 1 more variable: magnitude_acceleration <dbl>

```

Optional exercise (++) - Join datasets

```
data_crane_with_measures <- full_join(data_crane, data_crane_additional, by='event_id')
```

Optional exercise (+++) - gather all acceleration data

```
data_crane_long <- data_crane %>%
  filter(!is.na(acceleration_raw_x) &
         !is.na(acceleration_raw_y) &
         !is.na(acceleration_raw_z)) %>%
  gather(acceleration_direction, # column name for headers
         acceleration_speed, # column name for values
         acceleration_raw_x, acceleration_raw_y, acceleration_raw_z # columns to gather
        )

head(data_crane_long)

## # A tibble: 6 x 41
##   event_id visible timestamp           location_long location_lat
##   <dbl>     <lgl>   <dttm>              <dbl>          <dbl>
## 1 3.80e9  TRUE    2017-10-15 00:05:36      43.4          54.6
## 2 3.80e9  TRUE    2017-10-15 00:20:06      43.4          54.6
## 3 3.80e9  TRUE    2017-10-15 00:34:54      43.4          54.6
## 4 3.80e9  TRUE    2017-10-15 00:50:00      43.4          54.6
## 5 3.80e9  TRUE    2017-10-15 01:04:24      43.4          54.6
## 6 3.80e9  TRUE    2017-10-15 01:19:00      43.4          54.6
## # ... with 36 more variables: bar_barometric_height <dbl>,
## #   battery_charge_percent <dbl>, battery_charging_current <dbl>,
## #   eobs_activity <lgl>, eobs_activity_samples <lgl>,
## #   eobs_battery_voltage <dbl>, eobs_fix_battery_voltage <dbl>,
## #   eobs_horizontal_accuracy_estimate <dbl>, eobs_key_bin_checksum <dbl>,
## #   eobs_speed_accuracy_estimate <dbl>, eobs_start_timestamp <chr>,
## #   eobs_status <chr>, eobs_temperature <dbl>, eobs_type_of_fix <dbl>,
## #   eobs_used_time_to_get_fix <dbl>, external_temperature <dbl>,
## #   gps_hdop <dbl>, gps_satellite_count <dbl>, gps_time_to_fix <dbl>,
## #   ground_speed <dbl>, heading <dbl>, height_above_ellipsoid <dbl>,
## #   height_above_msl <dbl>, import_marked_outlier <lgl>,
## #   gls_light_level <dbl>, mag_magnetic_field_raw_x <dbl>,
## #   mag_magnetic_field_raw_y <dbl>, mag_magnetic_field_raw_z <dbl>,
## #   orn_transmission_protocol <chr>, tag_voltage <dbl>, sensor_type <chr>,
## #   individual_taxon_canonical_name <chr>,
## #   individual_local_identifier <chr>, study_name <chr>,
## #   acceleration_direction <chr>, acceleration_speed <dbl>
```