Problem Sets Prior to Test 1

*Only turn in problems that are **not** bracketed.* Bracketed problems are additional problems you can look at. Round brackets indicate problems that may help you with problems that are assigned; square brackets are additional problems on material that you should know, but you are not required to write up solutions; curly brackets are truly optional and may contain extra nuggets that you will not be required to know but may be interested in.

Additional assignments will be filled in over time.

notation	meaning
unbracketed	assigned problem – turn these in for grading
()	helper/warm-up problem
[]	additional problems (you are responsible for content, but don't turn them in)
{}	covers optional material

PS	Due	Source	Problems
0	Wed 2/1	handout	**1** Fill out personal information form. **2** Visit the course web page • http://www.calvin.edu/~rpruim/courses/s341/S17/ **3** Login to RStudio at http://rstudio.calvin.edu You should have recieved an email letting you know how to login and how to change your password.
1	Wed 2/1	Rethinking 2	(2M4) cards **2M5** cards **2M7** cards
2	Mon 2/6	Rethinking 2	**2E1–2E3** conditional probability **2H1–2H4** pandas
3	Wed 2/8	Rethinking 2 Rethinking 3	(2M1) grid **2M2** grid **2M3** earth or moon (3E1–3E7) posterior samples **3M1–3M3** globe [3M5]
4	~~Mon 2/13~~ Wed 2/15	Rethinking 3 Rethinking 4	**3H1–3H5** births **4E1–4E2** describing models **4M1** simulating prior **4M2** map See the code below for 3H1–3H5
5	Fri 2/17	Rethinking 4	**4E4–4E5** **4M3–4M6**
6	Mon 2/20	Rethinking 4	**4H1** **4H2** Howell kids **4H3** log-kg
7	Fri 2/24	Rethinking 5	(5E1) **5E2–3** creating models **5M4** LDS **5H1–2** foxes See the code below for some hints for Problem 5M4

Some Code for PS 4

I'm not particulary fond of how the author sets things up in 3H1–3H5. Here's a better way:

```
# load the birth1 and birth2 data vectors
data(homeworkch3, package = "rethinking")
# put them into a data frame
Birth <- data_frame(
  first = birth1,
  second = birth2
)
# tally up the counts
tally( ~ first + second, data = Birth, margins = TRUE)
```

```
##        second
## first     0   1 Total
##   0      10  39    49
##   1      30  21    51
##   Total  40  60   100
```

```
# another way to summarize:
Birth %>%
  # group by family type
  group_by(first, second) %>%
  summarise(
    # how many families of this type
    families = n(),
    # total boys in such families
    boys = sum(first + second),
    # total girls in such families
    girls = sum(2 - first - second)
  )
```

```
## Source: local data frame [4 x 5]
## Groups: first [?]
##
##   first second families  boys girls
##   <dbl>  <dbl>    <int> <dbl> <dbl>
## 1     0      0       10     0    20
## 2     0      1       39    39    39
## 3     1      0       30    30    30
## 4     1      1       21    42     0
```

Take a look at the code I've posted online for a way to use something other than 0's and 1's. In any case, the summaries above are all you need to do the problems.

Some Code for PS 7

Here is one way to get the data on Mormons (from Wikipedia).

```
require(rvest)
# doing this a crazy way to make it fit on the page
url <-
  paste0("https://en.wikipedia.org/wiki/",
         "The_Church_of_Jesus_Christ_of_Latter-day_Saints_",
         "membership_statistics_(United_States)"
  )
```

```
tables <- html_nodes(read_html(url), "table")
MormonsRaw <- html_table(tables[2], fill = TRUE)[[1]]
```

The numbers are formatted with commas in them, but we can use the following code to clean that up.

```
require(readr)
Mormons <-
  MormonsRaw %>%
  rename(LDS.prop = LDS) %>%
  mutate(
    Membership = parse_number(Membership),
    Population = parse_number(Population) / 1e6,  # in millions
    LDS.prop   = parse_number(LDS.prop)
    ) %>%
  select(1:4)  # keep only first four columns
```

Finally, we merge this in with the other data.

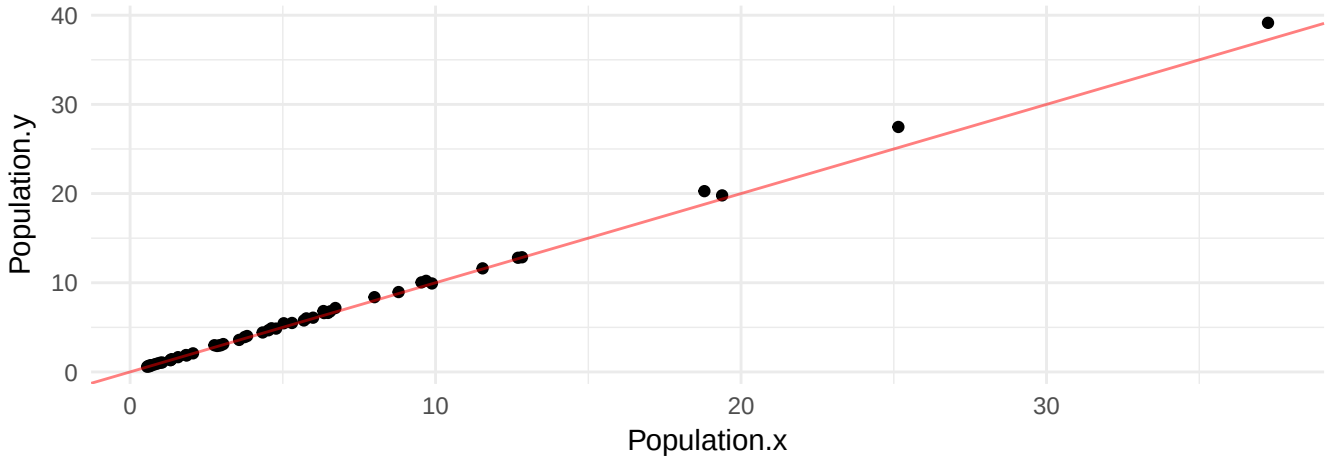
```
data(WaffleDivorce)
Divorce <-
  WaffleDivorce %>% rename(State = Location) %>%
  left_join(Mormons, by = "State")

## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining character vector and factor,
coercing into character vector

glimpse(Divorce)

## Observations: 50
## Variables: 16
## $ State             <chr> "Alabama", "Alaska", "Arizona", "Arkansas", ...
## $ Loc               <fctr> AL, AK, AZ, AR, CA, CO, CT, DE, DC, FL, GA,...
## $ Population.x      <dbl> 4.78, 0.71, 6.33, 2.92, 37.25, 5.03, 3.57, 0...
## $ MedianAgeMarriage <dbl> 25.3, 25.2, 25.8, 24.3, 26.8, 25.7, 27.6, 26...
## $ Marriage          <dbl> 20.2, 26.0, 20.3, 26.4, 19.1, 23.5, 17.1, 23...
## $ Marriage.SE       <dbl> 1.27, 2.93, 0.98, 1.70, 0.39, 1.24, 1.06, 2....
## $ Divorce           <dbl> 12.7, 12.5, 10.8, 13.5, 8.0, 11.6, 6.7, 8.9,...
## $ Divorce.SE        <dbl> 0.79, 2.05, 0.74, 1.22, 0.24, 0.94, 0.77, 1....
## $ WaffleHouses      <int> 128, 0, 18, 41, 0, 11, 0, 3, 0, 133, 381, 0,...
## $ South             <int> 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0,...
## $ Slaves1860        <int> 435080, 0, 0, 111115, 0, 0, 0, 1798, 0, 6174...
## $ Population1860    <int> 964201, 0, 0, 435450, 379994, 34277, 460147,...
## $ PropSlaves1860    <dbl> 4.5e-01, 0.0e+00, 0.0e+00, 2.6e-01, 0.0e+00,...
## $ Membership        <dbl> 36874, 33649, 418959, 30447, 773762, 151580,...
## $ Population.y      <dbl> 4.858979, 0.738432, 6.828065, 2.978204, 39.1...
## $ LDS.prop          <dbl> 0.76, 4.56, 6.14, 1.02, 1.98, 2.78, 0.44, 0....

require(statisticalModeling)  # for gf_ functions
# quick check that the two data sets give similar populations for each state
gf_point(Population.y ~ Population.x, data = Divorce) +
  geom_abline(intercept = 0, slope = 1, colour = "red", alpha = 0.5)
```


40
30
20
10
0
Population.y
0
10
20
30
Population.x