

Can the day a flower blooms be predicted? The law of the flowering plants.

Unit 7 Lecture 3

Jonathan Auerbach
STAT 489 Pre-Cap Prof Development
jauerba@gmu.edu



November 23, 2021

Learning Objectives

After this lecture, you will be able to:

1. State the law of the flowering plants and explain how Quetelet derived it.
2. Use the `rnoa` package to reproduce Quetelet's findings with weather data from the Global Historical Climatology Network.
3. Use the `rnpn` package to replicate Quetelet's findings with more recent data from the USA National Phenology Network.
4. Describe how the USA National Phenology Network uses the bloom dates of lilacs to monitor the start of spring.

These slides use the following R packages

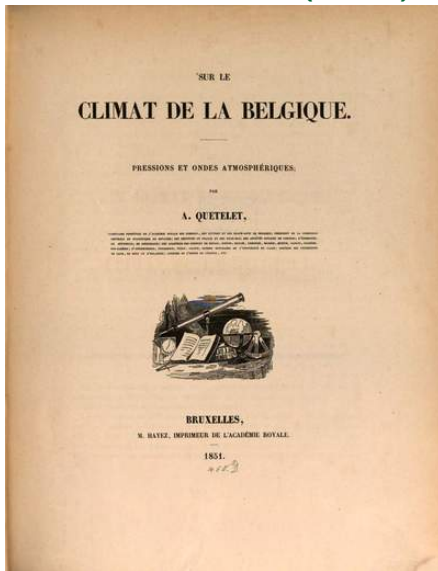
Setup:

```
library("knitr")  
library("tidyverse")  
library("rnoaa")  
library("rnpn")  
theme_set(theme_bw(base_size = 20))
```

Can the day a flower blooms be predicted?

- ▶ Anthophiles have recorded the dates flowers bloom for centuries.
 - ▷ e.g. written records of cherry trees go back as far as 812 A.D. in Japan and peach and plum trees as far as 1308 A.D. in China.
 - ▷ Systematic records begin with *Indicators of Spring* (Marsham, 1736)
- ▶ Adolph Quetelet recorded the bloom dates of hundreds of different plants between 1833 and 1852 at the Brussels Observatory.
 - ▷ Quetelet concluded that a plant flowers when exposed to a specific quantity of heat, measured in degrees of Celsius squared.
 - ▶ e.g. the lilac blooms when the sum of the mean daily temperature squared (following the last frost) exceeds 4264°C^2 .
- ▶ Quetelet communicated this law in his *Letters* (Number 33, 1849) and in his reporting *On the climate of Belgium* (1857).
 - ▷ Modern scientists use variations of the law to plan harvests, manage pests, and identify ecosystems stressed by climate change.

Quetelet and *On the climate of Belgium* (1851)



Source: https://en.wikipedia.org/wiki/Adolphe_Quetelet#/media/File:Adolphe_Qu%C3%A9telet_by_Joseph-Arnold_Demannez.jpg
<https://gallica.bnf.fr/ark:/12148/bpt6k95028d>

Bloom dates at Brussels Observatory 1839-1852

108

SUR LE CLIMAT DE LA BELGIQUE.

TABLEAU n° 15 (suite).

Floraison des plantes.

NOMS DES PLANTES.	1839-1844.	1845.	1846.	1847.	1848.	1849.	1850.	1851.	1852.	1845-1852.	écarts moyennes.
<i>Robinia pseudo-acacia</i>	29 mai.	12 juin.	4 juin.	30 mai.	17 mai.	1 juin.	7 juin.	15 juin.	14 juin.	4 juin.	2 juin.
<i>Rosa centifolia</i>	23 mai.	11 juin.	22 mai.	1 juin.	22 mai.	5 juin.	6 juin.	9 juin.	15 juin.	5 juin.	30 mai.
<i>Rubus idaeus</i>	8 mai.	2 juin.	5 mai.	23 mai.	7 mai.	15 mai.	24 mai.	24 mai.	25 mai.	19 mai.	15 mai.
<i>Sambucus racemosa</i>	17 avril.	1 mai.	27 mars.	2 mai.	5 avril.	14 avril.	17 avril.	21 avril.	24 avril.	17 avril.	17 avril.
<i>Id. nigra</i>	22 mai.	12 juin.	31 mai.	3 juin.	14 mai.	24 mai.	7 juin.	8 juin.	6 juin.	31 mai.	28 mai.
<i>Saxifraga cressifolia</i>	7 avril.	27 avril.	21 mars.	18 avril.	5 avril.	12 avril.	14 avril.	19 avril.	26 avril.	14 avril.	12 avril.
<i>Id. umbrosa</i>	9 mai.	28 mai.	15 avril.	15 mai.	5 mai.	5 mai.	18 mai.	9 mai.	15 mai.	10 mai.	10 mai.
<i>Scabiosa purpurea</i>	9 juillet.	"	8 août. ?	"	18 juin.	25 juin.	"	"	"	7 juillet.	8 juillet.
<i>Sedum acre</i>	6 juin.	19 juin.	15 juin.	21 juin.	18 juin.	25 juin.	26 juin.	"	"	21 juin.	16 juin.
<i>Id. album</i>	27 juin.	22 juin.	20 juin.	25 juin.	27 juin.	25 juin.	25 juin.	26 juin.	29 juin.	25 juin.	26 juin.
<i>Senecio jacobaeus</i>	12 juillet.	"	7 juillet.	"	9 juillet.	8 juillet.	"	"	"	8 juillet.	9 juillet.
<i>Sorbus aucuparia</i>	28 avril.	15 mai.	16 avril.	15 mai.	29 avril.	10 mai.	9 mai.	21 mai.	10 mai.	7 mai.	4 mai.
<i>Spiraea hypericifolia</i>	2 mai.	"	24 avril.	15 mai.	5 mai.	14 mai.	24 mai.	"	16 mai.	11 mai.	8 mai.
<i>Id. serotifolia</i>	12 juin.	17 juin.	"	5 juillet.	7 juin.	"	"	"	5 juillet.	24 juin.	20 juin.
<i>Id. salicifolia</i>	6 juin.	17 juin.	9 juin.	24 juin.	29 mai.	12 juin.	9 juin.	30 juin.	20 juin.	16 juin.	15 juin.
<i>Staphylea pinnata</i>	26 avril.	10 mai.	27 avril.	12 mai.	19 avril.	1 mai.	6 mai.	7 mai.	11 mai.	4 mai.	1 mai.
<i>Symphytum asperinum</i>	1 mai.	20 mai.	5 mai.	24 mai.	1 mai.	5 mai.	4 mai.	11 mai.	9 mai.	10 mai.	7 mai.
<i>Syringa vulgaris</i>	27 avril.	15 mai.	12 avril.	9 mai.	21 avril.	2 mai.	30 avril.	1 mai.	12 mai.	1 mai.	30 avril.
<i>Id. persica</i>	28 avril.	17 mai.	18 avril.	15 mai.	21 avril.	5 mai.	9 mai.	7 mai.	15 mai.	6 mai.	3 mai.
<i>Tiarella cordifolia</i>	26 avril.	7 mai.	15 avril.	15 mai.	"	5 mai.	25 avril.	1 mai.	10 mai.	5 mai.	1 mai.
<i>Tilia europaea</i>	5 juin.	17 juin.	6 juin.	15 mai.	6 juin.	12 juin.	14 juin.	25 juin.	20 juin.	11 juin.	9 juin.
<i>Tradescantia virginica</i>	27 mai.	14 juin.	9 juin.	8 juin.	25 mai.	4 juin.	1 juin.	10 juin.	4 juin.	6 juin.	5 juin.
<i>Trifolium pratense</i>	8 mai.	27 mai.	7 mai.	25 mai.	15 mai.	16 mai.	"	"	29 mai.	19 mai.	15 mai.
<i>Trollius europaeus</i>	5 mai.	15 mai.	24 avril.	15 mai.	21 avril.	5 mai.	9 mai.	9 mai.	15 mai.	6 mai.	5 mai.
<i>Tulipa Gesseri</i>	26 avril.	1 mai.	15 avril.	9 mai.	21 avril.	29 avril.	27 avril.	17 avril.	18 avril.	6 avril.	25 avril.
<i>Ulinus compestris</i>	20 mars.	7 avril.	4 février.	22 mars.	25 mars.	15 mars.	1 avril.	"	25 mars.	18 mars.	19 mars.
<i>Valeriana rubra</i>	21 mai.	12 juin.	10 mai.	31 mai.	16 mai.	25 mai.	1 juin.	27 mai.	23 mai.	20 mai.	21 mai.
<i>Verbascum plicatum</i>	8 mai.	18 juin.	19 mai.	21 mai.	12 mai.	24 mai.	"	"	"	24 mai.	19 mai.
<i>Veronica tectorum</i>	15 mai.	8 juin.	18 mai.	26 mai.	16 mai.	28 mai.	30 mai.	"	9 mai.	24 mai.	21 mai.
<i>Id. incana</i>	21 juin.	7 juillet.	22 juin.	25 juin.	12 juin.	17 juin.	27 juin.	2 juillet.	5 juillet.	20 juin.	24 juin.
<i>Viburnum opulus</i> , fl. pl.	10 mai.	28 mai.	1 mai.	19 mai.	9 mai.	15 mai.	10 mai.	14 mai.	14 mai.	15 mai.	15 mai.
<i>Vicia minor</i>	17 mars.	15 avril.	23 février.	5 avril.	23 mars.	28 février.	28 février.	20 mars.	20 mars.	17 mars.	17 mars.
<i>Viola odorata</i>	18 mars.	4 avril.	27 février.	22 mars.	18 mars.	15 mars.	10 mars.	14 mars.	17 février.	15 mars.	15 mars.
<i>Vitis vinifera</i>	21 juin.	6 juillet.	17 juin.	20 juin.	16 juin.	25 juin.	24 juin.	"	8 juillet.	25 juin.	24 juin.
<i>Walsteinia goeides</i>	1 avril.	21 avril.	5 mars.	10 avril.	31 mars.	8 mars.	4 avril.	28 mars.	28 mars.	29 mars.	30 mars.
<i>Yucca filamentosa</i>	2 juillet.	"	22 juin.	15 juillet.	6 juillet.	12 juillet.	9 juillet.	26 juillet.	14 juillet.	10 juillet.	7 juillet.

Lilac bloom dates, Brussels Observatory 1839-1852

```
lilac <- # Common Lilac, Syringa vulgaris
  tibble(month = c("May", "April", "April", "April",
                  "April", "April", "May", "April", "May",
                  "April", "May", "April", "May", "May"),
         day   = c(10, 28, 24, 28, 20, 25, 13,
                  12, 9, 21, 2, 30, 1, 12),
         year  = 1839:1852,
         date  = as.Date(paste(month, day, year),
                        format = "%B %d %Y"),
         doy   = parse_number(format(date, "%j")))

lilac %>% slice(1:3) %>% kable(align = "c")
```

month	day	year	date	doy
May	10	1839	1839-05-10	130
April	28	1840	1840-04-28	119
April	24	1841	1841-04-24	114

Daily temperature, Brussels Observatory 1839-1852

```
temp <- ghcnv_search(stationid = "BE000006447",
                    var = c("tmax", "tmin"),
                    date_min = "1839-01-01",
                    date_max = "1852-12-31") %>%
  reduce(left_join) %>%
  transmute(year = parse_number(format(date, "%Y")),
            date, tmax = tmax / 10, tmin = tmin / 10,
            temp = (tmax + tmin) / 2)

temp %>% slice(1:4) %>% select(year, date, temp) %>%
  kable(align = "c", col.names =
        c("year", "date", "midrange temperature °C"))
```

year	date	midrange temperature °C
1839	1839-01-01	2.75
1839	1839-01-02	3.55
1839	1839-01-03	4.50
1839	1839-01-04	4.90

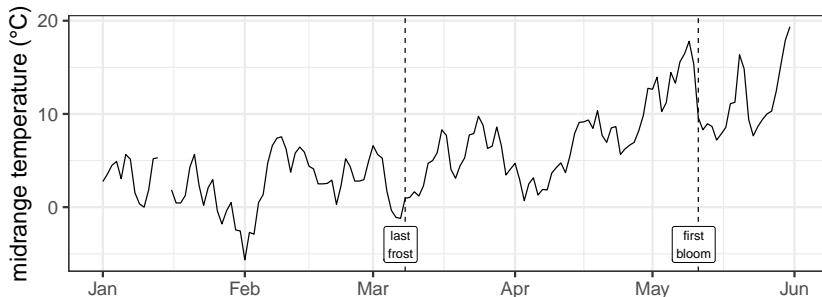
Assume frost if previous day's max temperature < 0

```
doy_last_frost <- function(temp, doy_max) {  
  dof <- which(temp[1:doy_max] < 0)  
  if(length(dof) == 0) {1} else {max(dof) + 1}  
  
bloom_day <- lilac %>% filter(year == 1839) %>%  
  pull(doy) + as.Date("1839-01-01")  
  
frost_day <- temp %>% filter(year == 1839) %>%  
  pull(tmax) %>%  
  doy_last_frost(100) + as.Date("1839-01-01")  
  
tibble(`last frost date` = frost_day,  
       `bloom date` = bloom_day) %>%  
kable(align = "c")
```

last frost date	bloom date
1839-03-08	1839-05-11

e.g. lilacs in 1839 exposed from March 8 to May 11

```
temp %>% filter(date < as.Date("1839-06-01")) %>%  
  ggplot() + aes(date, temp) + geom_line() +  
  labs(x = "", y = "midrange temperature (°C)") +  
  geom_vline(xintercept = frost_day, linetype = 2) +  
  geom_vline(xintercept = bloom_day, linetype = 2) +  
  annotate("label", y = -4,  
         label = c("last\nfrost", "first\nbloom"),  
         x = as.Date(c("1839-03-07", "1839-05-10")))
```



Quetelet's findings are reproducible...

```
quetelet <- temp %>% group_by(year) %>% nest() %>%  
  left_join(lilac) %>% mutate(threshold = map(data, ~  
    sum((pmax(.$temp, 0)[ # n.b. temperature < 0 set to 0  
      (doy_last_frost(.$tmax, doym) + 1):doy])^2,  
      na.rm = TRUE))) %>% unnest(threshold) %>% ungroup()  
  
quetelet %>%  
  summarize(law = 4264,  
            est = mean(threshold),  
            se = sd(threshold)/sqrt(n()),  
            ci = str_c("[", round(est - 2 * se), ", ",  
              round(est + 2 * se), "]")) %>%  
  kable(dig = 0, align = "c",  
        col.names = c("Quetelet's law", "estimate",  
                       "standard error", "confidence interval"))
```

Quetelet's law	estimate	standard error	confidence interval
4264	4263	197	[3868, 4657]

... predictions accurate within a week on average...

```
querelet %>%
  mutate(pred = map(data, ~
    which.max(
      cumsum(ifelse(is.na(.$temp), 0, pmax(.$temp, 0)^2)) -
      sum(pmax(.$temp, 0)[
        1:(doy_last_frost(.$tmax, 150))]^2,
        na.rm = TRUE) > 4264))) %>%
  unnest(pred) %>% ungroup() %>%
  summarize(
    mae = str_c(mean(abs(doy - pred)), " days"),
    rmse = str_c(round(sqrt(mean((doy - pred)^2))),
      " days")) %>%
  kable(align = "c",
    col.names = c("mean absolute error (MAE)",
      "root mean squared error (RMSE)"))
```

mean absolute error (MAE)	root mean squared error (RMSE)
---------------------------	--------------------------------

5 days

6 days

... and without temperature, mean error 40% higher

```
quietelet %>%
  summarize(
    pred = mean(doy),
    mae = str_c(round(mean(abs(doy - pred))), " days"),
    rmse = str_c(round(sqrt(mean((doy - pred)^2))),
                 " days")) %>%
  select(mae, rmse) %>%
  kable(align = "c",
        col.names = c("mean absolute error (MAE)",
                      "root mean squared error (RMSE)"))
```

mean absolute error (MAE)	root mean squared error (RMSE)
---------------------------	--------------------------------

7 days

9 days

Quetelet's law was inspired by physics

- ▶ Quetelet was not the first to explore the relationship between temperature and bloom dates.
 - ▷ René Réaumur noticed the relationship in 1735, a century earlier.
 - ▷ But Quetelet was the first to systematically study the relationship across a wide variety of species and empirically derive the amount of heat plants needed to bloom, which varies by species.

- ▶ Quetelet believed that, as in physics, universal laws govern biological and social phenomenon.
 - ▷ In fact, Quetelet's law is a direct application of Newton's second law of motion.
 - ▷ Newton's second law states that acceleration is proportional to force: An object initially at rest and subject to a constant force travels a distance proportional to time squared.
 - ▷ Quetelet simply substituted temperature for time.

Distance object travels if subject to constant force

Let $d(t)$ denote the distance an object travels after time t . Let $v(t) = d'(t)$ denote its speed and $a(t) = v'(t)$ its acceleration. If acceleration is constant, i.e. $a(t) = c$,

$$v(t) = \int_0^t a(s) ds = \int_0^t c ds = ct$$

and

$$d(t) = \int_0^t v(s) ds = \int_0^t cs ds = \frac{c}{2}t^2$$

Quetelet imagined plants experience time in temperature and bloom after “traveling” distance d_* . If a plant is exposed to temperature t_i on day $i = 1, 2, \dots$, then the bloom date, n_* , is the first day $\sum_{i=1}^{n_*} \frac{c}{2}t_i^2 \geq d_*$.

Multiplying both sides of the inequality by $\frac{2}{c}$, yields the law: the bloom is the first day, n_* , that $\sum_{i=1}^{n_*} t_i^2 \geq \frac{2}{c}d_*$.

Do flowers actually obey Newton's laws?

- ▶ The derivation of laws like the law of the flowering plants was popular in the nineteenth century.
 - ▷ But the relationship between the force of temperature and the force of gravity is coincidental.
 - ▷ No known biological mechanism justifies Quetelet's application of Newton's law.
 - ▷ Today, the law of the flowering plants is considered a heuristic, or rule of thumb, that approximates complicated biological mechanisms
- ▶ Like Quetelet, scientists model plants as experiencing time in temperature instead of calendar time.
 - ▷ These temperature units are typically called "growing degree days."
 - ▷ Scientists often find that plants may only be sensitive to temperatures in specific ranges or "modified growing degree days."
 - ▷ Although modern statistical methods can greatly improve the accuracy of predictions, laws like Quetelet's remain popular because they are simple and easily replicated.

Quetelet's findings replicable with USA NPN data...

```
usa_npn %>% # See Appendix for creating usa_npn
  group_by(rownames(usa_npn)) %>%
  mutate(threshold = map(temp, ~ sum((pmax(.$temp, 0)[
    (doy_last_frost(.$tmax, doy) + 1):(doy-1)]))^2,
    na.rm = TRUE))) %>% unnest(threshold) %>%
  group_by(year) %>% # adjust for dependence within year
  summarize(threshold = mean(threshold)) %>%
  summarize(law = 4264,
            est = mean(threshold),
            se = sd(threshold) / sqrt(n()),
            ci = str_c("[", round(est - 2 * se), ", ",
            round(est + 2 * se), "]")) %>%
  kable(dig = 0, align = "c",
        col.names = c("Quetelet's law", "estimate",
                      "standard error", "confidence interval"))
```

Quetelet's law	estimate	standard error	confidence interval
4264	4416	112	[4192, 4641]

... predictions accurate within 2 weeks on average...

```
usa_npn <- usa_npn %>%
  mutate(pred = map(temp, ~ which.max(
    cumsum(iffelse(is.na(. $temp), 0, pmax(. $temp, 0)^2)) -
    sum(pmax(. $temp, 0)[1:(doy_last_frost(. $tmax, 150))]^2,
      na.rm = TRUE) > 4264))) %>% unnest(pred) %>% ungroup()

usa_npn %>% summarize(
  mae = str_c(round(mean(abs(doy - 1 - pred))), " days"),
  rmse = str_c(round(sqrt(mean((doy - 1 - pred)^2))),
    " days")) %>%
  kable(align = "c",
    col.names = c("mean absolute error (MAE)",
      "root mean squared error (RMSE)"))
```

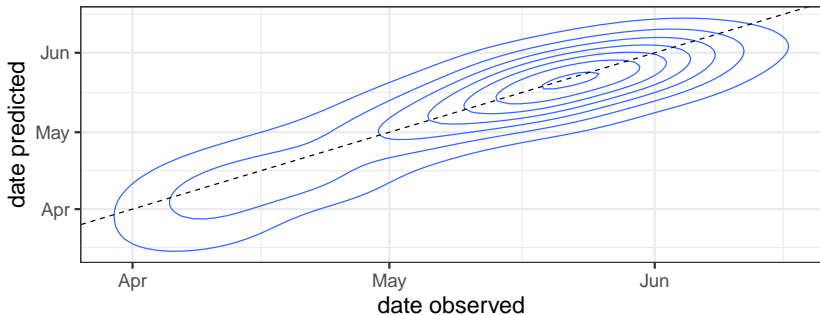
mean absolute error (MAE)	root mean squared error (RMSE)
---------------------------	--------------------------------

10 days

15 days

... predictions accurate within 2 weeks on average...

```
usa_npn %>% mutate(doy = first_yes_doy) %>%  
  unnest(pred) %>% ungroup() %>%  
  ggplot() + geom_density2d() +  
  aes(as.Date("2020-01-01") + doym,   
      as.Date("2020-01-01") + pred) +  
  geom_abline(intercept = 0, slope = 1, linetype = 2) +  
  labs(x = "date observed",  
       y = "date predicted") +  
  theme(legend.position = "none")
```



... and without temperature, mean error 2x higher

```
usa_npn %>%  
  mutate(doy = first_yes_doy) %>%  
  ungroup() %>%  
  summarize(  
    pred = mean(quetelet$doy),  
    mae = str_c(round(mean(abs(doy - pred))),  
                " days"),  
    rmse = str_c(round(sqrt(mean((doy - pred)^2))),  
                 " days")) %>%  
  select(mae, rmse) %>%  
  kable(align = "c",  
        col.names = c("mean absolute error (MAE)",  
                       "root mean squared error (RMSE)"))
```

mean absolute error (MAE)	root mean squared error (RMSE)
---------------------------	--------------------------------

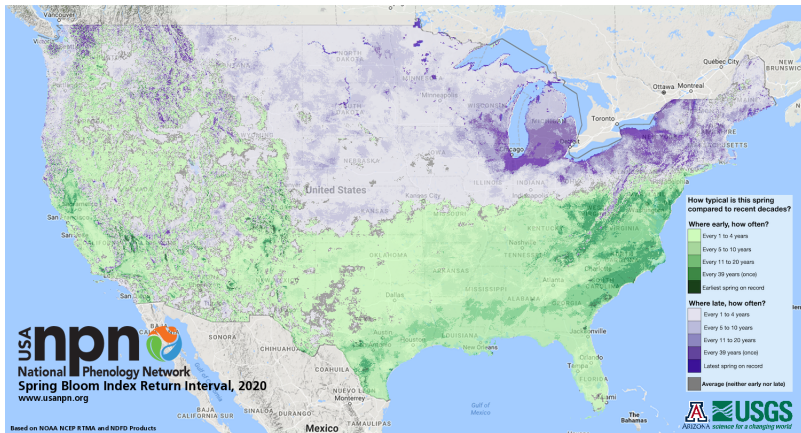
21 days

24 days

Quetelet's greatest legacy: advocate and mentor

- ▶ Quetelet founded an international network for “observations of the periodical phenomena,” in addition to numerous statistical societies and publications.
 - ▷ Quetelet's network of 80 stations collected observations throughout Europe from 1841 until 1872.
 - ▷ The data were collected with help from Charles Morren—who later coined the term phenology, the name of the field that studies biological lifecycle events like the timing of flower blooms.
- ▶ Though Quetelet's emphasis on discovering the universal laws he believed govern social and biological phenomenon has not endured...
 - ▷ ... scientists continue to use laws like Quetelet's to plan harvests, manage pests, and identify ecosystems stressed by climate change.
 - ▷ e.g. the USA National Phenology Network calculates the spring bloom index using the day lilacs bloom across the United States.
 - ▷ Bloom dates that are atypically early or atypically late can indicate environments impacted by changing climate.

The return interval measures how typical spring is



Source: <https://www.usanpn.org/files/nbn/maps/six-bloom-return-interval-2020.png>

References

1. Caprio, J.M., Schwartz M.D. and the USA National Phenology Network. 2013. Lilac and Honeysuckle Data for the United States, 1956-2013. <https://www.usanpn.org/home>
2. Demarée, Gaston and This Rutishauser. From “Periodical Observations” to “Anthochronology” and “Phenology”. *International Journal of Biometeorology* 55.6 (2011): 753-761.
3. Quetelet, Adolphe. *Sur le climat de la Belgique*. vol. 2. 1857.
4. Quetelet, Adolphe. *Letters addressed to HRH the grand duke of Saxe Coburg and Gotha: On the theory of probabilities, as applied to the moral and political sciences*. C. & E. Layton, 1849.
5. Stigler, Stephen. *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press, 1986.

Appendix: buds (L), open (C), and full flowers (R)

Quetelet marks the day of flowering as the day on which “the first corolla opens and shows the stamina.” That event is closest to the phenophase “open flowers,” depicted in the center image below.



Source: <https://www.usanpn.org/nn/TrackALilac>

Appendix: Preparing USA NPN Data

```
# 1. download lilac data using `rnpn`
usa_npn <-
  npn_download_individual_phenometrics(
    request_source = "Jonathan Auerbach",
    year = 1900:2050,
    species_ids = 36,          # common lilac, Syringa vulgaris
    phenophase_ids = c(77, 412)) # first bloom

# 2. limit analysis to sites that report more than 25 times
site_ids <- usa_npn %>% group_by(site_id) %>%
  summarize(n = n()) %>% filter(n > 25) %>% pull(site_id)

usa_npn <- usa_npn %>% filter(site_id %in% site_ids)

# 3. find nearest weather stations for each site
locations <- usa_npn %>% group_by(site_id) %>%
  summarize(latitude = first(latitude),
             longitude = first(longitude))
```

Appendix: Preparing USA NPN Data (2)

```
stations <- ghcn_d_stations() %>%
  filter(first_year <= min(usa_npn$first_year),
         last_year >= max(usa_npn$first_year),
         state != "") %>%
  group_by(id, latitude, longitude, state) %>%
  summarize(temp_flag = # remove stations without tmax/tmin
            sum(element %in% c("TMIN", "TMAX"))) %>%
  filter(temp_flag == 2) %>% ungroup()

dist <- function(x, y = stations %>%
                select(latitude, longitude))
  stations$id[which.min(sqrt((x[1] - y[,1])^2 +
                           (x[2] - y[,2])^2)[,1])]

locations$station_id <-
  apply(locations, 1, function(x)
    dist(c(x["latitude"], x["longitude"])))
```

Appendix: Preparing USA NPN Data (3)

```
# 4. get weather data from nearest station using `rnoaa`
get_station_data <- function(station_id)
  ghcnd_search(stationid = station_id,
               var = c("tmin", "tmax"),
               date_min = "1956-01-01",
               date_max = "2011-12-31") %>%
  reduce(left_join, by = c("id", "date")) %>%
  select(id, date, tmin, tmax)

usa_npn <- locations %>%
  mutate(temp = map(station_id, get_station_data)) %>%
  right_join(usa_npn, by = "site_id") %>%
  group_by(rownames(usa_npn)) %>% mutate(
    temp = map(temp, ~ .x %>%
      filter(format(date, "%Y") == first_yes_year) %>%
      mutate(temp = (tmin + tmax) / 20)),
    num_obs = map(temp, ~ sum(format(.x$date, "%j") <= 150)),
    doy = first_yes_doy, year = first_yes_year) %>%
  unnest(num_obs) %>% filter(num_obs == 150)
```