

# Michelson Speed-of-light Measurements

Michaela Fox 2025-02-06

- [Grading Rubric](#)
  - [Individual](#)
  - [Submission](#)
    - [q1](#) Re-create the following table (from Michelson (1880), pg. 139) using `df_michelson` and `dplyr`. Note that your values *will not* match those of Michelson *exactly*; why might this be?
    - [q2](#) Create a new variable `VelocityVacuum` with the `+$92$` km/s adjustment to `Velocity`. Assign this new dataframe to `df_q2`.
    - [q3](#) Compare Michelson’s speed of light estimate against the modern speed of light value. Is Michelson’s estimate of the error (his uncertainty) greater or less than the true error?
    - [q4](#) Inspect the following plot with the `Real` Michelson data and `Simulated` data from a probability model. Document the similarities and differences between the data under *observe* below.
    - [q5](#) You have access to a few other variables. Construct a **at least three** visualizations of `VelocityVacuum` against these other factors. Are there other patterns in the data that might help explain the difference between Michelson’s estimate and `LIGHTSPEED_VACUUM`?
  - [Bibliography](#)

*Purpose:* When studying physical problems, there is an important distinction between *error* and *uncertainty*. The primary purpose of this challenge is to dip our toes into these factors by analyzing a real dataset.

*Reading:* [Experimental Determination of the Velocity of Light](#) (Optional)

## Grading Rubric {#grading-rubric}

Unlike exercises, **challenges will be graded**. The following rubrics define how you will be graded, both on an individual and team basis.

### Individual {#individual}

Category	Needs Improvement	Satisfactory
Effort	Some task <b>q</b> ’s left unattempted	All task <b>q</b> ’s attempted
Observed	Did not document observations, or observations incorrect	Documented correct observations based on analysis
Supported	Some observations not clearly supported by analysis	All observations clearly supported by analysis (table, graph, etc.)

Category	Needs Improvement	Satisfactory
Assessed	Observations include claims not supported by the data, or reflect a level of certainty not warranted by the data	Observations are appropriately qualified by the quality & relevance of the data and (in)conclusiveness of the support
Specified	Uses the phrase "more data are necessary" without clarification	Any statement that "more data are necessary" specifies which <i>specific</i> data are needed to answer what <i>specific</i> question
Code Styled	Violations of the <a href="#">style guide</a> hinder readability	Code sufficiently close to the <a href="#">style guide</a>

## Submission {#submission}

Make sure to commit both the challenge report ([report.md](#) file) and supporting files ([report\\_files/](#) folder) when you are done! Then submit a link to Canvas. **Your Challenge submission is not complete without all files uploaded to GitHub.**

```
# Libraries
library(tidyverse)
library(google sheets4)

url <- "https://docs.google.com/spreadsheets/d/1av_SXn4j0-4Rk0mQFik3LLr-uf0YdA06i3ugE6n-Zdo/edit?usp=sharing"

# Parameters
LIGHTSPEED_VACUUM <- 299792.458 # Exact speed of light in a vacuum (km / s)
LIGHTSPEED_MICHELSON <- 299944.00 # Michelson's speed estimate (km / s)
LIGHTSPEED_PM <- 51 # Michelson error estimate (km / s)
```

**Background:** In 1879 Albert Michelson led an experimental campaign to measure the speed of light. His approach was a development upon the method of Foucault, and resulted in a new estimate of  $v_0 = 299944 \pm 51$  kilometers per second (in a vacuum). This is very close to the modern *exact* value of  $2.9979246 \times 10^8$ . In this challenge, you will analyze Michelson's original data, and explore some of the factors associated with his experiment.

I've already copied Michelson's data from his 1880 publication; the code chunk below will load these data from a public googlesheet.

**Aside:** The speed of light is *exact* (there is **zero error** in the value `LIGHTSPEED_VACUUM`) because the meter is actually *defined* in terms of the speed of light!

```
## Note: No need to edit this chunk!
gs4_deauth()
ss <- gs4_get(url)
df_michelson <-
  read_sheet(ss) %>%
```

```
select(Date, Distinctness, Temp, Velocity) %>%
mutate(Distinctness = as_factor(Distinctness))
```

```
## ✔ Reading from "michelson1879".
```

```
## ✔ Range 'Sheet1'.
```

```
df_michelson %>% glimpse()
```

```
## Rows: 100
## Columns: 4
## $ Date      <dtm> 1879-06-05, 1879-06-07, 1879-06-07, 1879-06-07, 1879-06-...
## $ Distinctness <fct> 3, 2, 2, 2, 2, 2, 3, 3, 3, 3, 2, 2, 2, 2, 2, 1, 3, 3, 2, ...
## $ Temp      <dbl> 76, 72, 72, 72, 72, 72, 83, 83, 83, 83, 83, 90, 90, 71, 7...
## $ Velocity   <dbl> 299850, 299740, 299900, 300070, 299930, 299850, 299950, 2...
```

Data dictionary:

- **Date**: Date of measurement
- **Distinctness**: Distinctness of measured images: 3 = good, 2 = fair, 1 = poor
- **Temp**: Ambient temperature (Fahrenheit)
- **Velocity**: Measured speed of light (km / s)

**q1** Re-create the following table (from Michelson (1880), pg. 139) using `df_michelson` and `dplyr`. Note that your values *will not* match those of Michelson *exactly*; why might this be?

Distinctness	n	MeanVelocity
3	46	299860
2	39	299860
1	15	299810

```
## TODO: Compute summaries
df_q1 <- df_michelson %>%
  group_by(Distinctness) %>%
  summarise(
    n = n(),
    MeanVelocity = mean(Velocity, na.rm = TRUE)
  )

df_q1 %>%
```

```
arrange(desc(Distinctness)) %>%
knitr::kable()
```

Distinctness	n	MeanVelocity
3	46	299861.7
2	39	299858.5
1	15	299808.0

**Observations:** - Write your observations here! - (Your response here) - Why might your table differ from Michelson’s? - (Your response here)

Compared to Michelson’s, the computed MeanVelocity seems to be rounded to the tens place, while Michelson’s is rounded to a whole number. This suggests that Michelson’s original value is the actual speed of light rounded to a whole number.

The **Velocity** values in the dataset are the speed of light *in air*; Michelson introduced a couple of adjustments to estimate the speed of light in a vacuum. In total, he added \$+92\$ km/s to his mean estimate for **VelocityVacuum** (from Michelson (1880), pg. 141). While the following isn’t fully rigorous (\$+92\$ km/s is based on the mean temperature), we’ll simply apply this correction to all the observations in the dataset.

**q2** Create a new variable **VelocityVacuum** with the \$+92\$ km/s adjustment to **Velocity**. Assign this new dataframe to **df\_q2**.

```
## TODO: Adjust the data, assign to df_q2
df_q2 <- df_michelson %>%
  mutate(VelocityVacuum = Velocity + 92)

df_q2
```

```
## # A tibble: 100 × 5
##   Date          Distinctness Temp Velocity VelocityVacuum
##   <dtm>         <fct>      <dbl>    <dbl>         <dbl>
## 1 1879-06-05 00:00:00 3         76    299850         299942
## 2 1879-06-07 00:00:00 2         72    299740         299832
## 3 1879-06-07 00:00:00 2         72    299900         299992
## 4 1879-06-07 00:00:00 2         72    300070         300162
## 5 1879-06-07 00:00:00 2         72    299930         300022
## 6 1879-06-07 00:00:00 2         72    299850         299942
## 7 1879-06-09 00:00:00 3         83    299950         300042
## 8 1879-06-09 00:00:00 3         83    299980         300072
## 9 1879-06-09 00:00:00 3         83    299980         300072
## 10 1879-06-09 00:00:00 3         83    299880         299972
## # i 90 more rows
```

As part of his study, Michelson assessed the various potential sources of error, and provided his best-guess for the error in his speed-of-light estimate. These values are provided in `LIGHTSPEED_MICHELSON`—his nominal estimate—and `LIGHTSPEED_PM`—plus/minus bounds on his estimate. Put differently, Michelson believed the true value of the speed-of-light probably lay between `LIGHTSPEED_MICHELSON - LIGHTSPEED_PM` and `LIGHTSPEED_MICHELSON + LIGHTSPEED_PM`.

Let's introduce some terminology:\$\$\$

- **Error** is the difference between a true value and an estimate of that value; for instance `LIGHTSPEED_VACUUM - LIGHTSPEED_MICHELSON`.
- **Uncertainty** is an analyst's *assessment* of the error.

Since a "true" value is often not known in practice, one generally does not know the error. The best they can do is quantify their degree of uncertainty. We will learn some means of quantifying uncertainty in this class, but for many real problems uncertainty includes some amount of human judgment. \$\$\$

**q3** Compare Michelson's speed of light estimate against the modern speed of light value. Is Michelson's estimate of the error (his uncertainty) greater or less than the true error?

```
## TODO: Compare Michelson's estimate and error against the true value
true_value <- LIGHTSPEED_MICHELSON - LIGHTSPEED_VACUUM
true_value
```

```
## [1] 151.542
```

```
LIGHTSPEED_PM
```

```
## [1] 51
```

```
if (LIGHTSPEED_PM > true_value) {
  print("Michelson overestimated")
} else if (LIGHTSPEED_PM < true_value) {
  print("Michelson underestimated")
}
```

```
## [1] "Michelson underestimated"
```

**Observations:** - Is Michelson's estimate of the error (his uncertainty) greater or less than the true error? Michelson's estimate of his error was less than the actual error, with his error being +- 51 and the real error

being +151.542.

- Make a quantitative comparison between Michelson's uncertainty and his error.

Michelson's uncertainty was about 3 times less than his actual error. This shows he was very confident in his answer but might have forgotten to account for a physical difference between his experimental setup and reality.

The following plot shows all of Michelson's data as a [control chart](#); this sort of plot is common in manufacturing, where it is used to help determine if a manufacturing process is under [statistical control](#). Each dot is one of Michelson's measurements, and the grey line connects the mean taken for each day. The same plot also shows simulated data using a probability model. We'll get into statistics later in the course; for now, let's focus on understanding what real and simulated data tend to look like.

**q4** Inspect the following plot with the **Real** Michelson data and **Simulated** data from a probability model. Document the similarities and differences between the data under *observe* below.

```
## Note: No need to edit this chunk!
## Calibrate simulated data
v_mean <-
  df_q2 %>%
  summarize(m = mean(VelocityVacuum)) %>%
  pull(m)
v_sd <-
  df_q2 %>%
  summarize(s = sd(VelocityVacuum)) %>%
  pull(s)

## Visualize
set.seed(101)
df_q2 %>%
  mutate(Simulated = rnorm(n(), mean = v_mean, sd = v_sd)) %>%
  rename(Real = VelocityVacuum) %>%
  pivot_longer(
    cols = c(Simulated, Real),
    names_to = "source",
    values_to = "velocity"
  ) %>%

  ggplot(aes(Date, velocity)) +
  geom_hline(
    yintercept = LIGHTSPEED_MICHELSON,
    linetype = "dotted"
  ) +
  geom_hline(
    yintercept = LIGHTSPEED_MICHELSON - LIGHTSPEED_PM,
    linetype = "dashed"
  ) +
  geom_hline(
    yintercept = LIGHTSPEED_MICHELSON + LIGHTSPEED_PM,
```

```

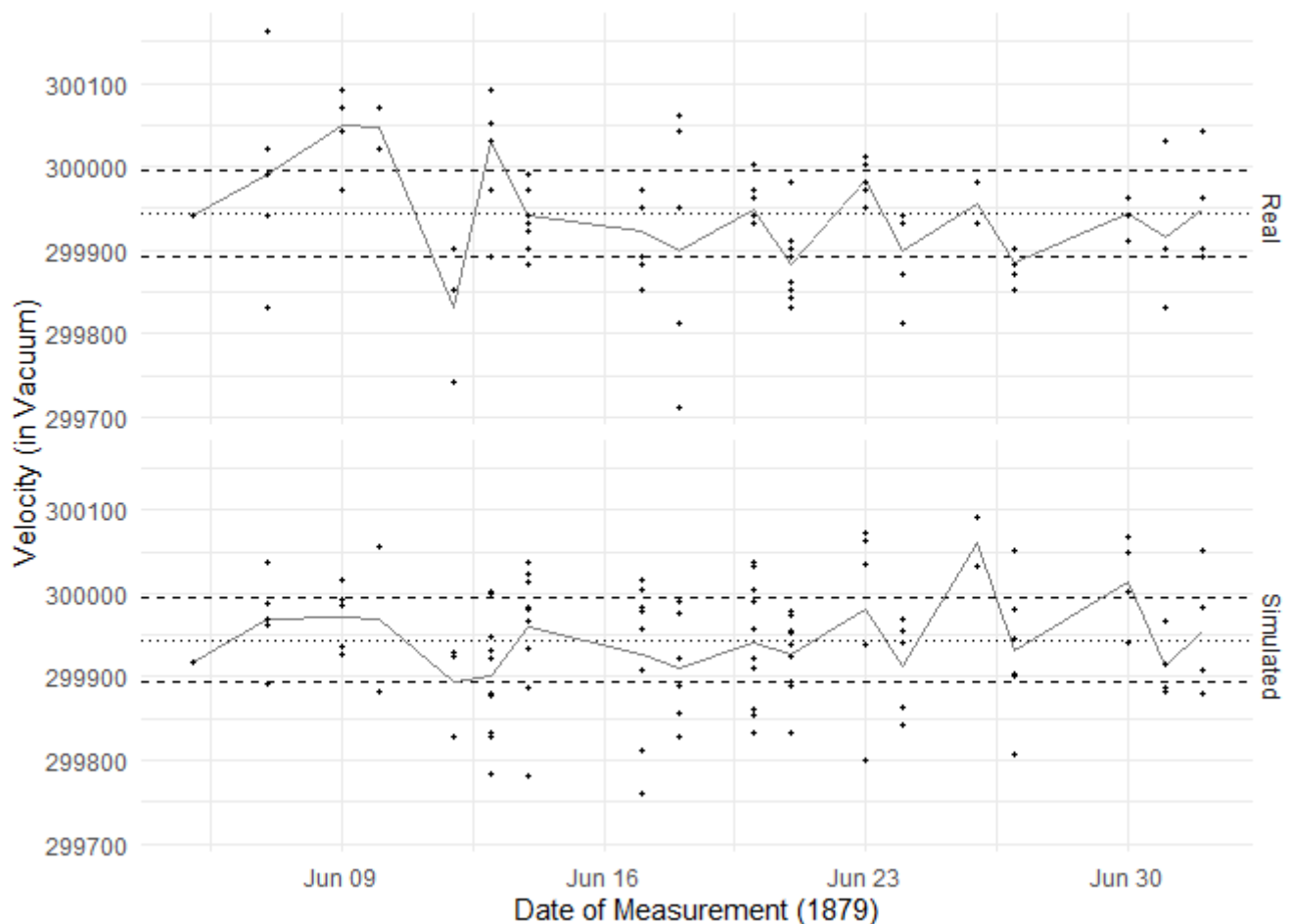
  linetype = "dashed"
) +

geom_line(
  data = . %>%
    group_by(Date, source) %>%
    summarize(velocity_mean = mean(velocity)),
  mapping = aes(y = velocity_mean),
  color = "grey50"
) +
geom_point(
  mapping = aes(y = velocity),
  size = 0.8
) +

facet_grid(source~.) +
theme_minimal() +
labs(
  x = "Date of Measurement (1879)",
  y = "Velocity (in Vacuum)"
)

```

## `summarise()` has grouped output by 'Date'. You can override using the  
## `groups` argument.

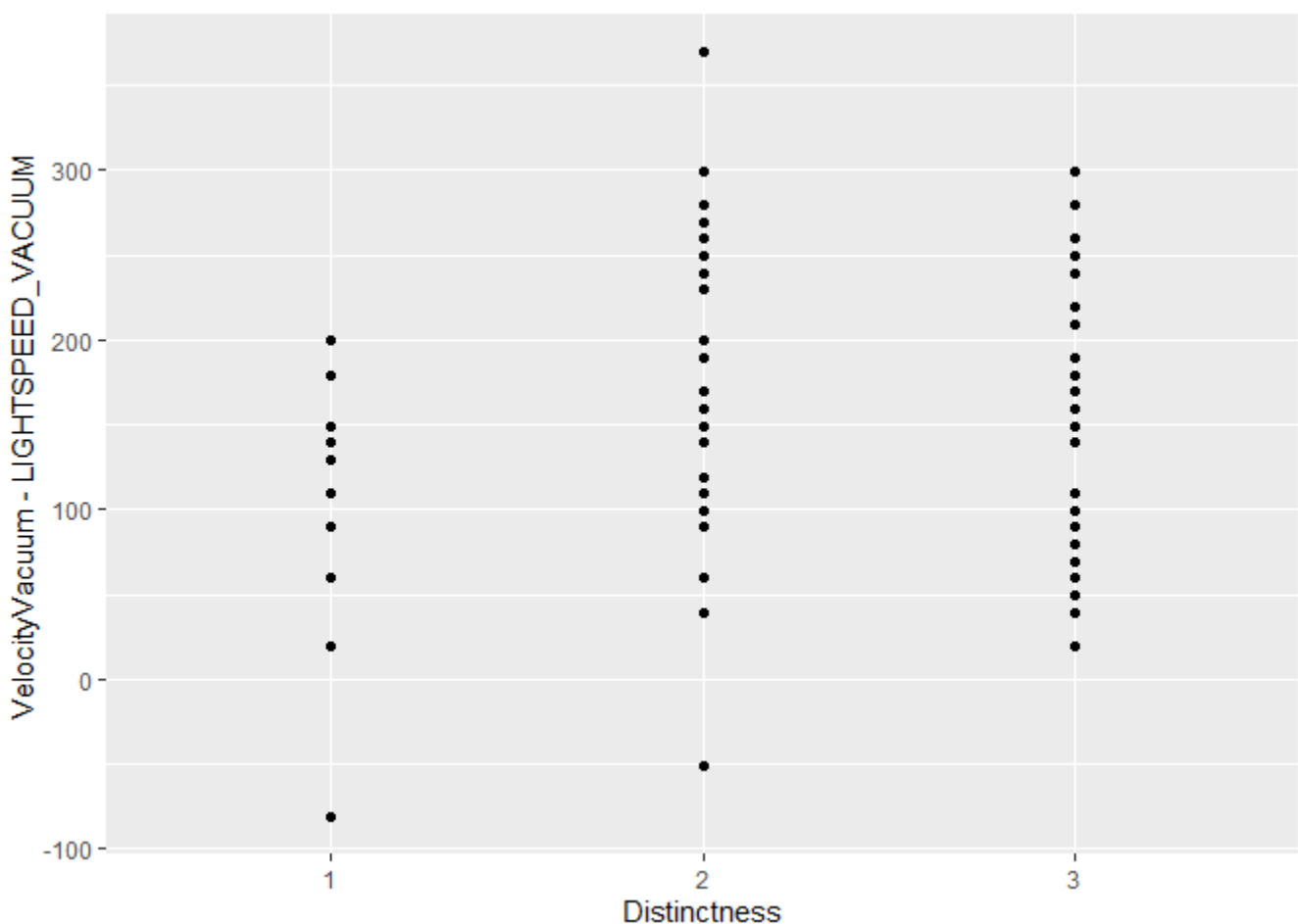


**Observations:** Similarities - The mean of both velocities stays inbetween 299800 and 300100 all days. Both means tend to increase and decrease at around the same times.

Differences - The real data has more extreme outliers. The real data tends to vary more from day to day.

**q5** You have access to a few other variables. Construct a **at least three** visualizations of **VelocityVacuum** against these other factors. Are there other patterns in the data that might help explain the difference between Michelson's estimate and **LIGHTSPEED\_VACUUM**?

```
# Difference between velocity and true value of lightspeed v. Distinctness
df_q2 %>%
  ggplot(aes(x = Distinctness, y = VelocityVacuum - LIGHTSPEED_VACUUM)) +
  geom_point()
```



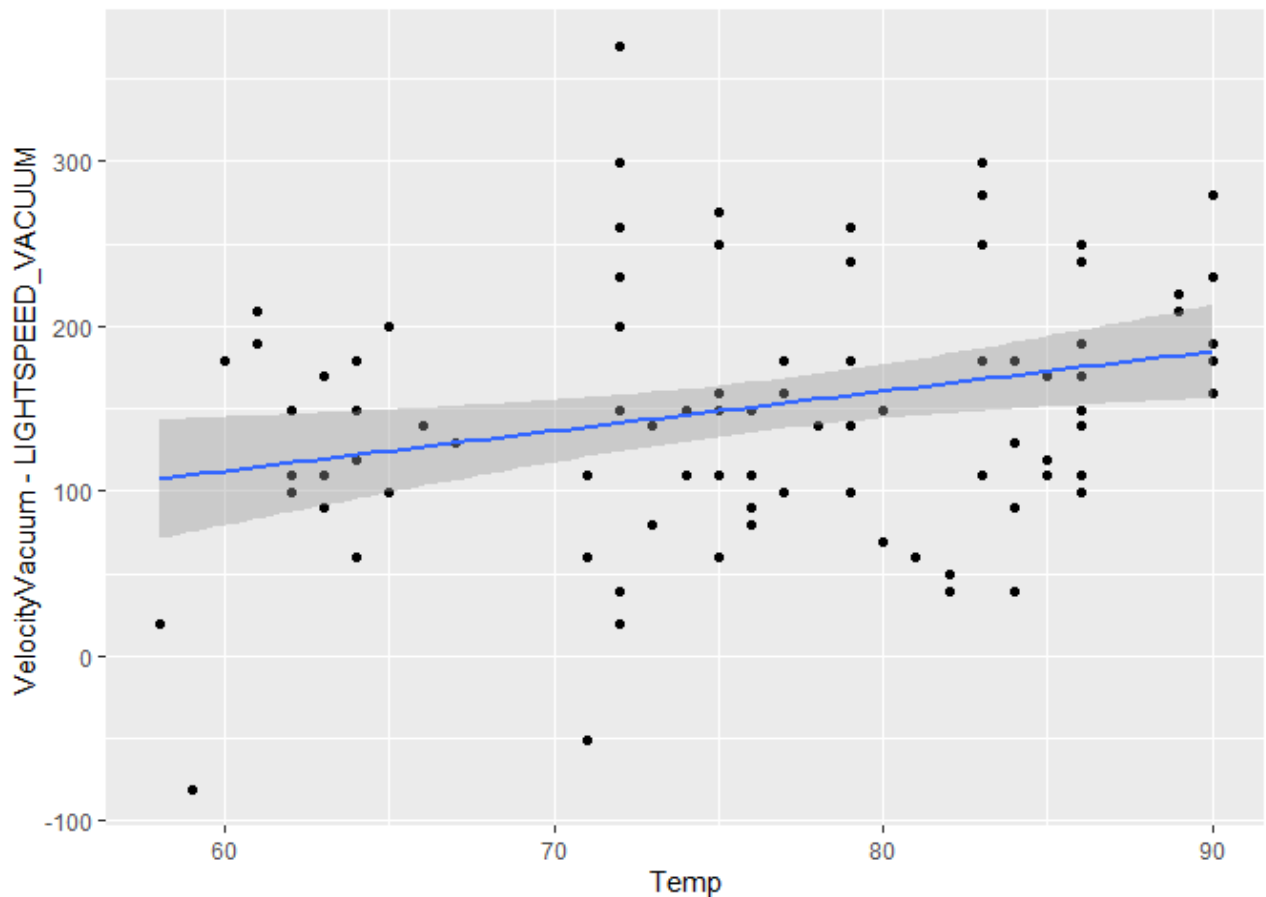
**Observations:**

- The spread of the difference is large for all distinctness values. This implies that distinctness was most likely not the cause of Michelson's error.
- The more distinct images actually deviate more from the true value, further supporting the observation that distinctness is not the cause of the error.
- There is no noticeable trend in the graph.



```
# Difference between velocity and true value of lightspeed v. Temperature
df_q2 %>%
  ggplot(aes(x = Temp, y = VelocityVacuum - LIGHTSPEED_VACUUM)) +
  geom_point() +
  geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

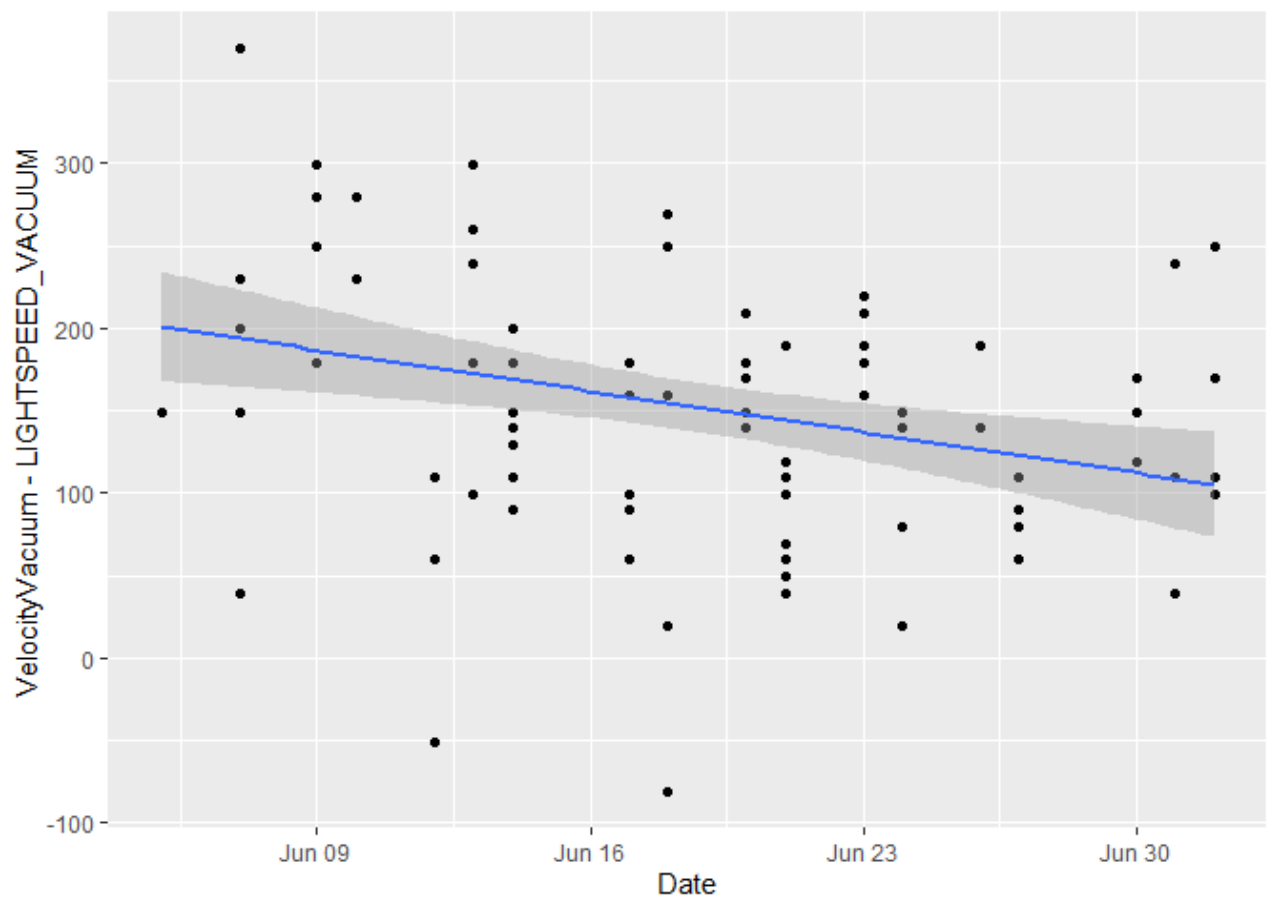


### Observations:

- While there is a slight correlation between temperature and the error in Michelson's estimate, the data is also very spread out.
- Temperature might have had a small influence on Michelson's error, but was probably not the main cause.
- Lower temperatures tended to have a smaller error than higher temperatures.

```
# Difference between velocity and true value of lightspeed v. Date
df_q2 %>%
  ggplot(aes(x = Date, y = VelocityVacuum - LIGHTSPEED_VACUUM)) +
  geom_point() +
  geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



### Observations:

- Michelson's error tended to decrease as time passed.
- Combining this knowledge with the fact that there was a slight correlation between temperature and error it can be assumed that Michelson's error might have something to do with time.
- However, like time, the data is still scattered, implying that there is some other reason for his error.

### Bibliography {#bibliography}

- \$1\$ Michelson, [Experimental Determination of the Velocity of Light](#) 1880)
- \$2\$ Henrion and Fischhoff, [Assessing Uncertainty in Physical Constants](#) 1986)
- \$3\$ BYU video about a [Fizeau-Foucault apparatus](#), similar to what Michelson used.