Full dataset analysis

First read in all of the relevant data.

```
all_ws_1 <-
  readInWebCDI(all_data_ws1_path) %>%
  select( #drop a bunch of columns that were screwing up the merge with prolific data
    -opt_out,
    -country,
    -sibling_boolean,
    -sibling_data,
    -sibling_count,
    -caregiver_other
  )
```

```
## Warning: Problem with `mutate()` input `Total Produced Percentile-sex`.
## i NAs introduced by coercion
## i Input `Total Produced Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) .

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Total Produced Percentile-both`.
## i NAs introduced by coercion
## i Input `Total Produced Percentile-both` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1)

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Complexity Percentile-sex`.
## i NAs introduced by coercion
## i Input `Complexity Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Combination Example 1`.
## i NAs introduced by coercion
## i Input `Combination Example 1` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Combination Example 2`.
## i NAs introduced by coercion
## i Input `Combination Example 2` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion
```

```
## Warning: Problem with `mutate()` input `Combination Example 3`.
## i NAs introduced by coercion
## i Input `Combination Example 3` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion

## Warning: Problem with `mutate()` input `other_languages`.
## i NAs introduced by coercion
## i Input `other_languages` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion

## Warning: Problem with `mutate()` input `language_from`.
## i NAs introduced by coercion
## i Input `language_from` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion
```

```
all_ws_2 <-
  readInWebCDI(all_data_ws2_path) %>%
  select( #drop a bunch of columns that were screwing up the merge with prolific data
    -opt_out,
    -country,
    -sibling_boolean,
    -sibling_data,
    -sibling_count,
    -caregiver_other
  )
```

```
## Warning: Problem with `mutate()` input `Total Produced Percentile-sex`.
## i NAs introduced by coercion
## i Input `Total Produced Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) .

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Total Produced Percentile-both`.
## i NAs introduced by coercion
## i Input `Total Produced Percentile-both` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1)

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Complexity Percentile-sex`.
## i NAs introduced by coercion
## i Input `Complexity Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Combination Example 1`.
## i NAs introduced by coercion
## i Input `Combination Example 1` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.
```

```
## Warning in ~as.numeric(.): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Combination Example 2`.
## i NAs introduced by coercion
## i Input `Combination Example 2` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Combination Example 3`.
## i NAs introduced by coercion
## i Input `Combination Example 3` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion

## Warning: Problem with `mutate()` input `other_languages`.
## i NAs introduced by coercion
## i Input `other_languages` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion

## Warning: Problem with `mutate()` input `language_from`.
## i NAs introduced by coercion
## i Input `language_from` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion
```

```
all_ws_raw <-
  all_ws_1 %>%
  bind_rows(all_ws_2) %>%
  mutate(completed = case_when(
    stringr::str_to_lower(completed) == "true" ~ TRUE,
    stringr::str_to_lower(completed) == "false" ~ FALSE
  ))

all_wg_raw <- readInWebCDI(all_data_wg_path)
```

```
## Warning: Problem with `mutate()` input `Phrases Percentile-sex`.
## i NAs introduced by coercion
## i Input `Phrases Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Words Understood Percentile-sex`.
## i NAs introduced by coercion
## i Input `Words Understood Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1)

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Words Understood Percentile-both`.
## i NAs introduced by coercion
## i Input `Words Understood Percentile-both` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1
```

```
## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Words Produced Percentile-sex`.
## i NAs introduced by coercion
## i Input `Words Produced Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) .

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Words Produced Percentile-both`.
## i NAs introduced by coercion
## i Input `Words Produced Percentile-both` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1)

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Early Gestures Percentile-sex`.
## i NAs introduced by coercion
## i Input `Early Gestures Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) .

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Early Gestures Percentile-both`.
## i NAs introduced by coercion
## i Input `Early Gestures Percentile-both` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1)

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Later Gestures Percentile-sex`.
## i NAs introduced by coercion
## i Input `Later Gestures Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) .

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Later Gestures Percentile-both`.
## i NAs introduced by coercion
## i Input `Later Gestures Percentile-both` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1)

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Total Gestures Percentile-sex`.
## i NAs introduced by coercion
## i Input `Total Gestures Percentile-sex` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) .

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion

## Warning: Problem with `mutate()` input `Total Gestures Percentile-both`.
## i NAs introduced by coercion
## i Input `Total Gestures Percentile-both` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1)

## Warning in eval_tidy(pair$rhs, env = default_env): NAs introduced by coercion
```

```
## Warning: Problem with `mutate()` input `other_languages`.
## i NAs introduced by coercion
## i Input `other_languages` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion

## Warning: Problem with `mutate()` input `language_from`.
## i NAs introduced by coercion
## i Input `language_from` is `(structure(function (..., .x = ..1, .y = ..2, . = ..1) ...`.

## Warning in ~as.numeric(.): NAs introduced by coercion
```

```
save(
  all_ws_raw,
  file = path(
    project_root,
    "data",
    "full_dataset",
    "unfiltered",
    "ws_unfiltered.RData"
  )
)

save(
  all_wg_raw,
  file = path(
    project_root,
    "data",
    "full_dataset",
    "unfiltered",
    "wg_unfiltered.RData"
  )
)
```

Filter out: multilingual exposure, illnesses, vision and hearing problems.

```
#original sample size of 2868
#WG
wg_filtered <-
  all_wg_raw %>%
  filter(repeat_num == "1") %>%
  filterBirthweight() %>%
  filterMultilingual() %>%
  filterIllnesses() %>%
  filterVision() %>%
  filterHearing() %>%
  getCompletionInterval() %>%
  getEthnicities() %>%
  getMaternalEd() %>%
  filter(completion_time >= min_completion_time) %>%
  filter_age_wg()
```

```
## Warning: Problem with `mutate()` input `maternal_ed`.
## i Unknown levels in `f`: Not reported
## i Input `maternal_ed` is `fct_relevel(...)`.

## Warning: Unknown levels in `f`: Not reported
```

```
wg_exclusion_n <- nrow(all_wg_raw) - nrow(wg_filtered)

save(
  wg_filtered,
  file = path(
    project_root,
    "data",
    "full_dataset",
    "filtered",
    "wg_filtered.RData"
  )
)

ws_filtered <-
  all_ws_raw %>%
  filter(repeat_num == "1") %>%
  filterBirthweight() %>%
  filterMultilingual() %>%
  filterIllnesses() %>%
  filterVision() %>%
  filterHearing() %>%
  getCompletionInterval() %>%
  getEthnicities() %>%
  getMaternalEd() %>%
  filter(completion_time >= min_completion_time) %>%
  filter_age_ws()

ws_exclusion_n <- nrow(all_ws_raw) - nrow(ws_filtered)

save(
  ws_filtered,
  file = path(
    project_root,
    "data",
    "full_dataset",
    "filtered",
    "ws_filtered.RData"
  )
)

total_n <- nrow(all_ws_raw) + nrow(all_wg_raw)

filtered_n <- nrow(ws_filtered) + nrow(wg_filtered)

nrow(all_wg_raw)
```

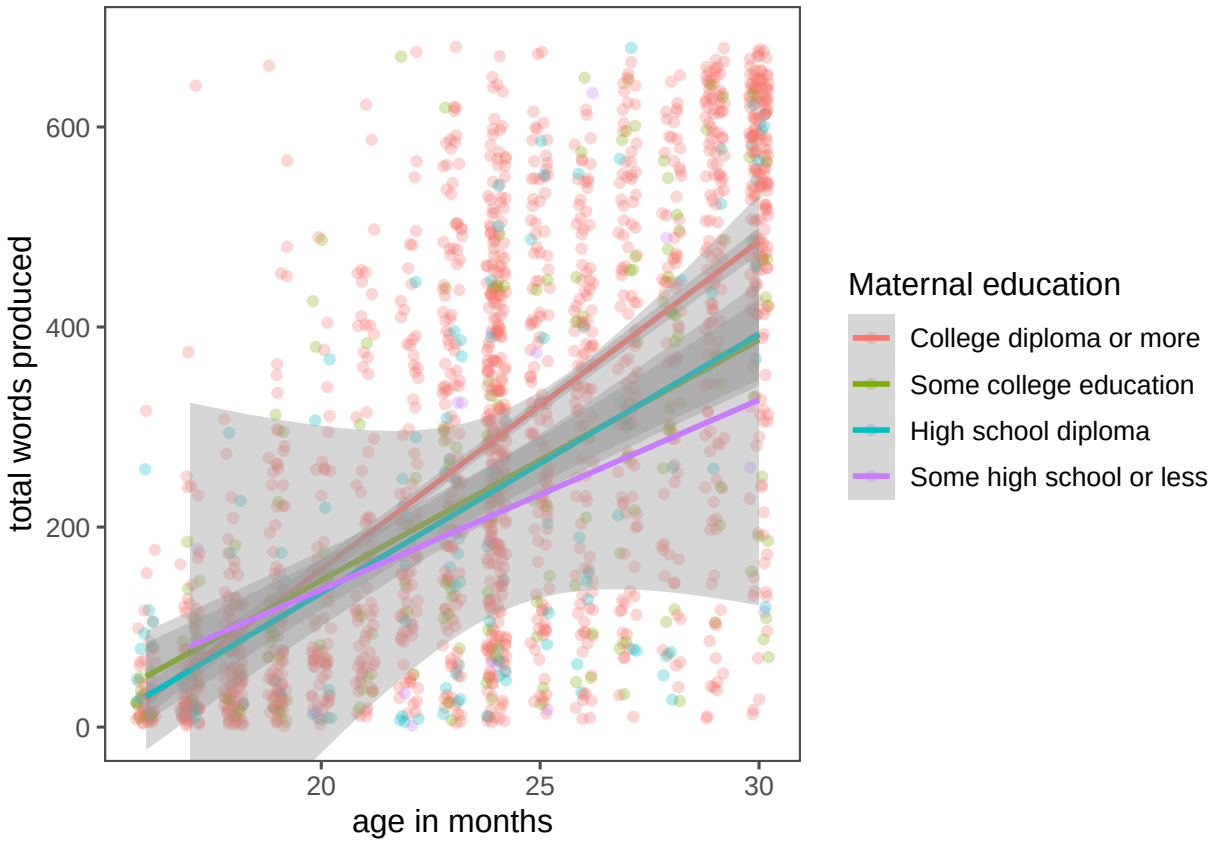
```
## [1] 2868
```

```
nrow(all_ws_raw)
```

```
## [1] 3594
```

```
#Comprehension and production measures
ws_filtered %>%
  filter(`Total Produced` < 688 & maternal_ed != "Not reported") %>%
  ggplot(aes(age, `Total Produced`, color = maternal_ed)) +
  ggthemes::theme_few() +
  geom_jitter(alpha = 0.3, width = 0.225) +
  coord_cartesian(ylim = c(0, 686)) +
  geom_smooth(method = "lm") +
  labs(
    x = "age in months",
    y = "total words produced",
    color = "Maternal education"
  )
```

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



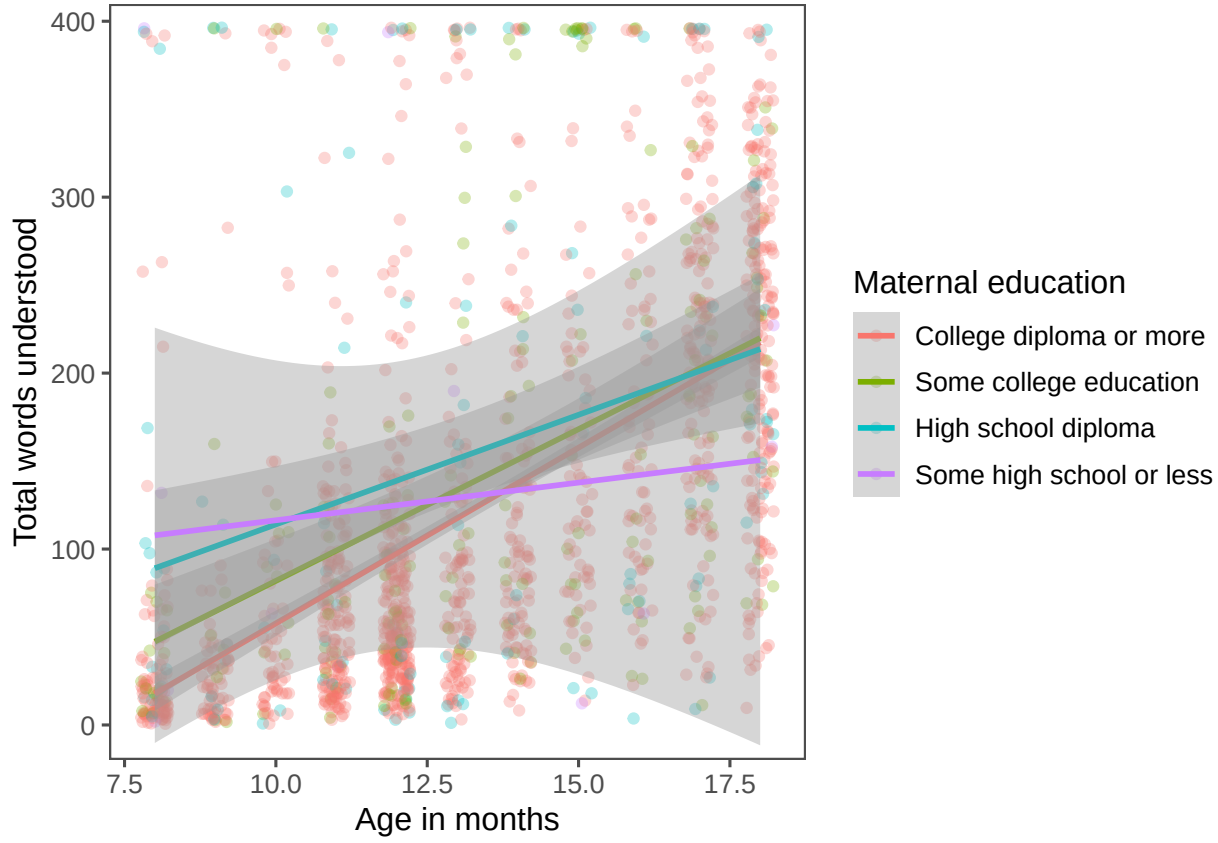

```
wg_filtered %>%
  filter(!is.na(maternal_ed)) %>%
  ggplot(aes(age, `Words Understood`, color = maternal_ed)) +
  ggthemes::theme_few() +
```

```
  geom_jitter(alpha = 0.3, width = 0.225) +
  geom_smooth(method = "lm") +
  coord_cartesian(ylim = c(0, 390)) +
  labs(
    x = "Age in months",
    y = "Total words understood",
    color = "Maternal education"
  )
```

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

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```



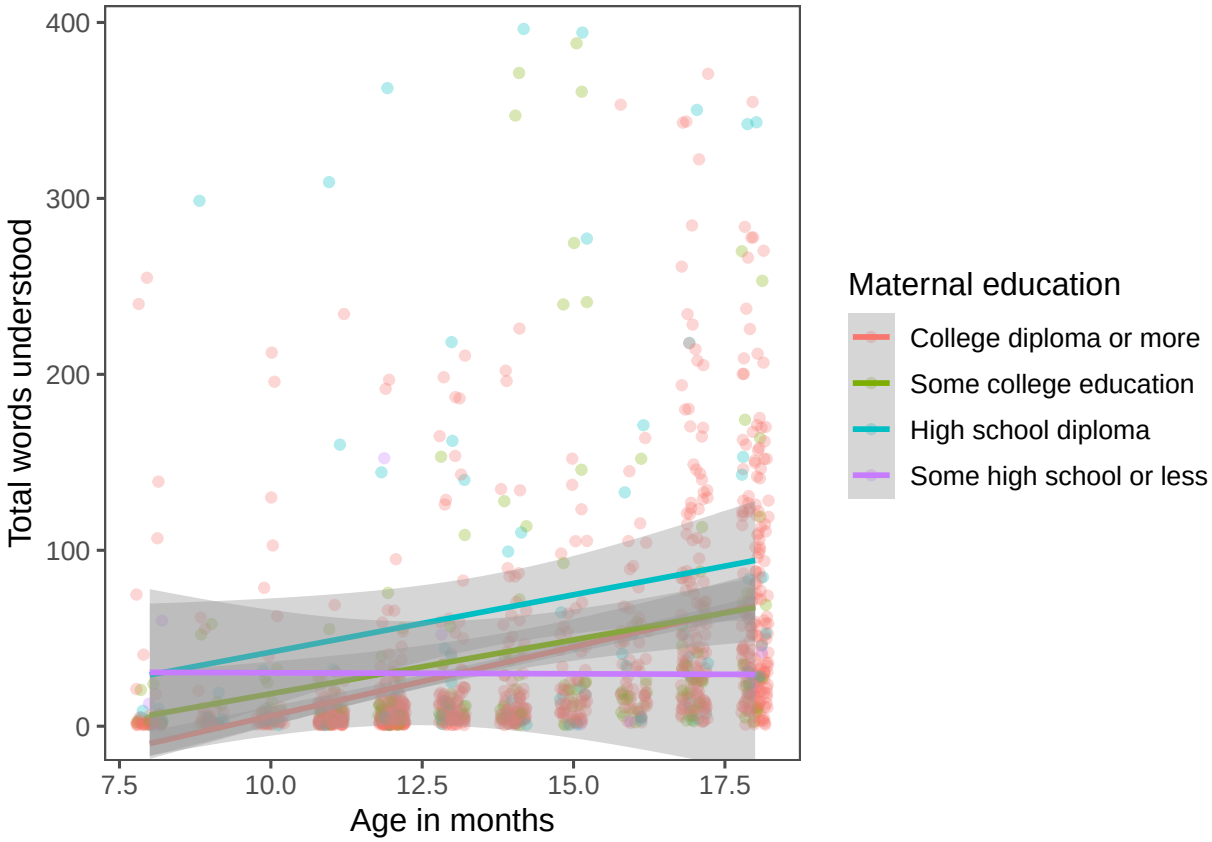

```
wg_filtered %>%
  filter(!is.na(maternal_ed)) %>%
  ggplot(aes(age, `Words Produced`, color = maternal_ed)) +
  ggthemes::theme_few() +
  geom_jitter(alpha = 0.3, width = 0.225) +
  geom_smooth(method = "lm") +
  coord_cartesian(ylim = c(0, 390)) +
  labs(
    x = "Age in months",
```

```
    y = "Total words understood",
    color = "Maternal education"
  )
```

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

```
## Warning: Removed 262 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 262 rows containing missing values (geom_point).
```



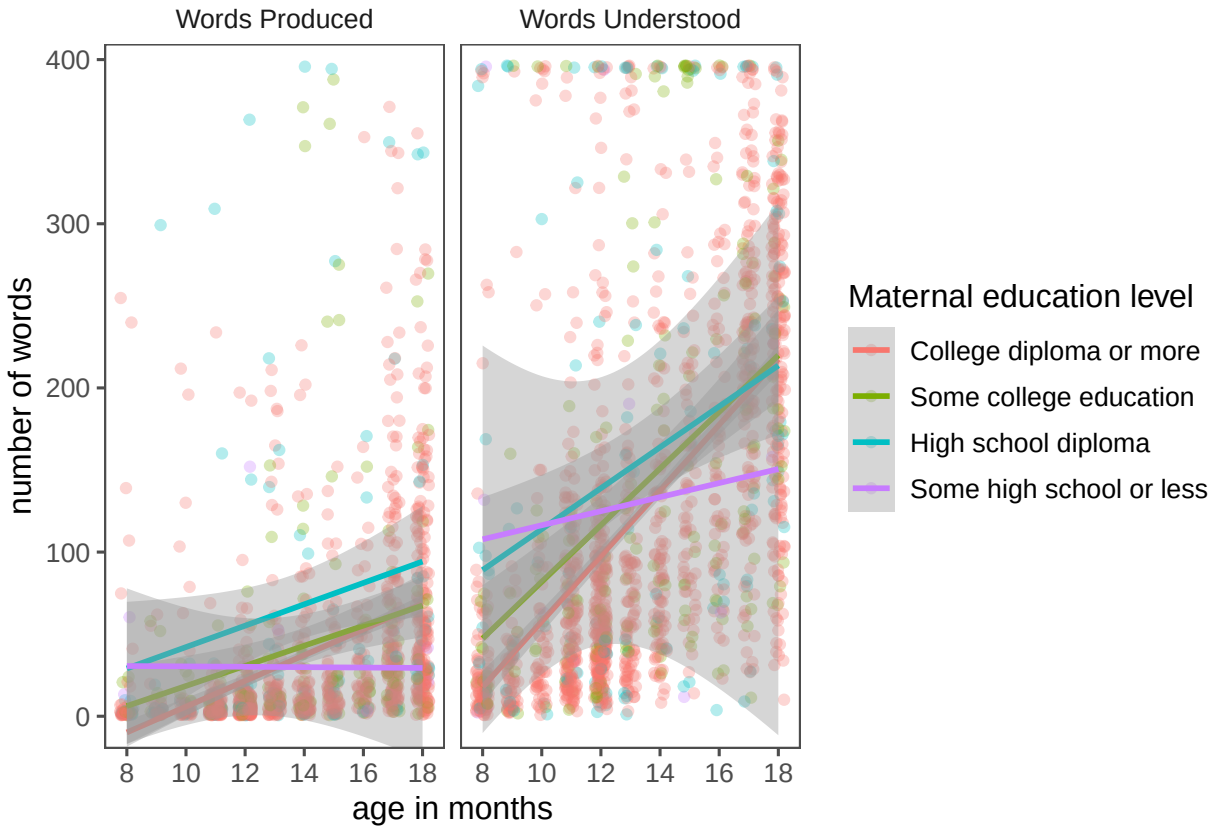

```
wg_filtered %>%
  filter(!is.na(maternal_ed)) %>%
  select(age, `Words Produced`, `Words Understood`, maternal_ed) %>%
  pivot_longer(
    cols = c("Words Produced", "Words Understood"),
    names_to = "measure",
    values_to = "words"
  ) %>%
  ggplot(aes(age, words, color = maternal_ed)) +
  facet_grid(~measure) +
  geom_jitter(alpha = 0.3, width = 0.225) +
  geom_smooth(method = "lm") +
  coord_cartesian(ylim = c(0, 390)) +
  ggthemes::theme_few() +
```

```
  labs(
    color = "Maternal education level",
    x = "age in months",
    y = "number of words"
  ) +
  scale_x_continuous(breaks = seq(from = 8, to = 18, by = 2))
```

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

```
## Warning: Removed 269 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 269 rows containing missing values (geom_point).
```



Gender analyses

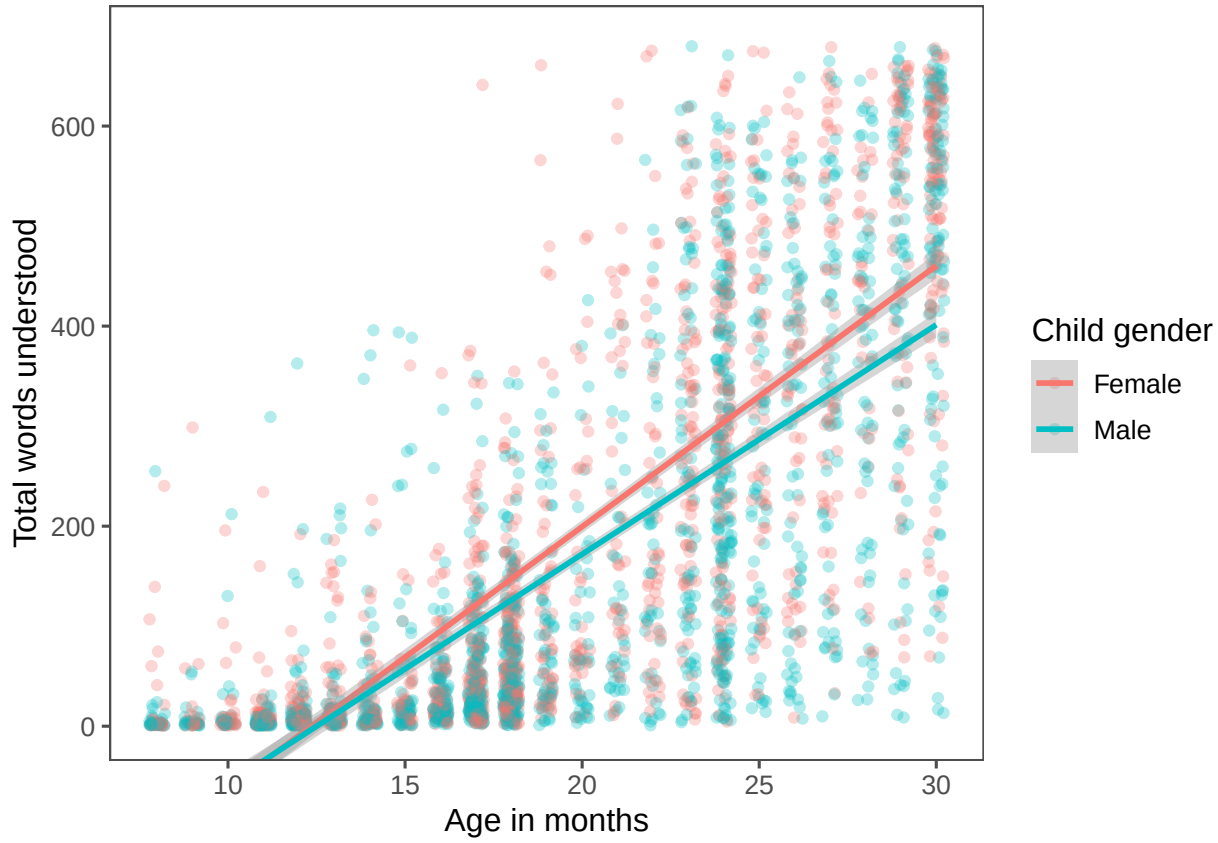
```
wg_filtered_gender <-
  wg_filtered %>%
  select(age, production = `Words Produced`, sex)

ws_filtered_gender <-
  ws_filtered %>%
  select(age, production = `Total Produced`, sex)

all_d_gender <-
```

```
  bind_rows(ws_filtered_gender, wg_filtered_gender)

all_d_gender %>%
  filter(sex != "Other") %>%
  ggplot(aes(age, production, color = sex)) +
  ggthemes::theme_few() +
  geom_jitter(alpha = 0.3, width = 0.225) +
  geom_smooth(method = "lm") +
  coord_cartesian(ylim = c(0, 686)) +
  labs(
    x = "Age in months",
    y = "Total words understood",
    color = "Child gender"
  )
```

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

```
## Warning: Removed 272 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 272 rows containing missing values (geom_point).
```




```
#Counting hispanic and latino heritage
wg_filtered %>%
  mutate(hispanic = !is.na(child_hispanic_latino) & child_hispanic_latino == TRUE) %>%
```

```
  count(hispanic) %>%
  mutate(prop = n / sum(n))
```

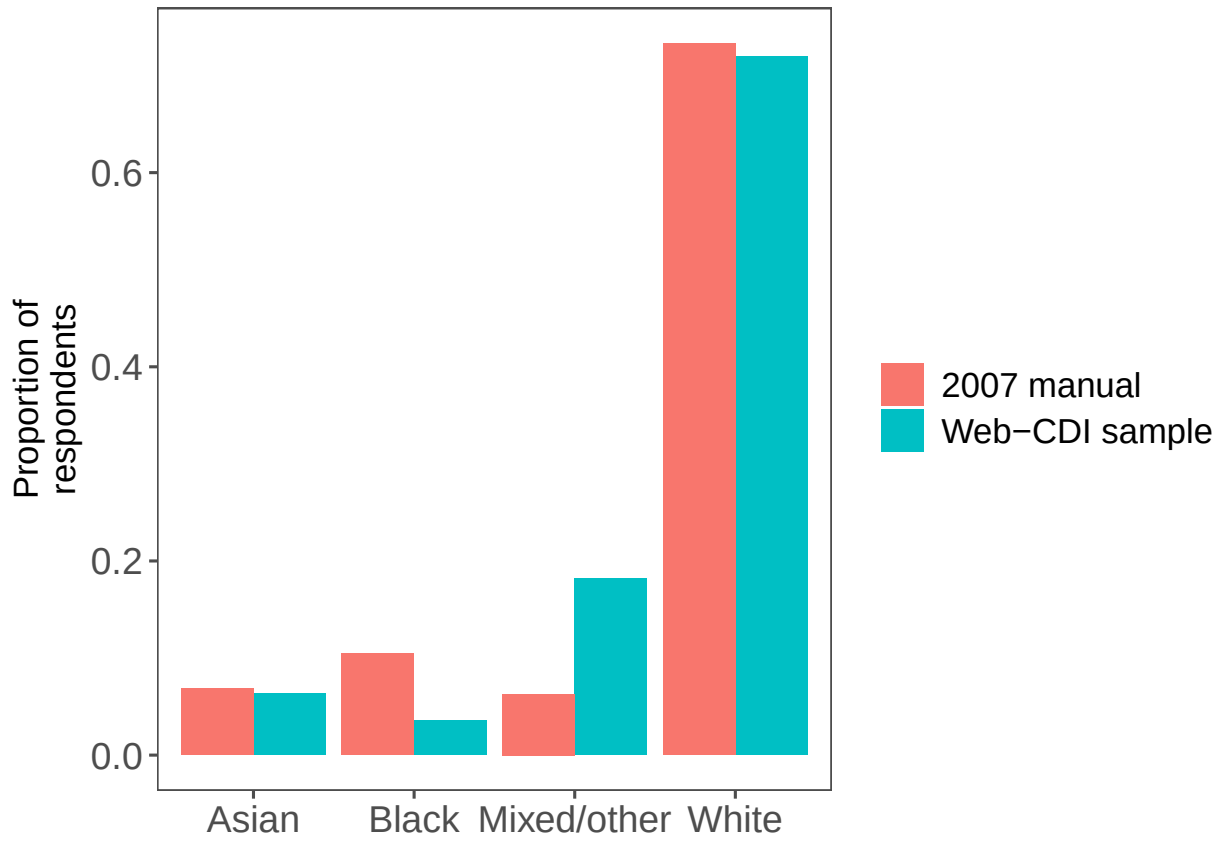
```
## # A tibble: 2 x 3
##   hispanic     n   prop
##   <lgl>    <int>  <dbl>
## 1 FALSE     1475 0.935
## 2 TRUE       102 0.0647
```

```
ws_filtered %>%
  mutate(hispanic = !is.na(child_hispanic_latino) & child_hispanic_latino == TRUE) %>%
  count(hispanic)%>%
  mutate(prop = n / sum(n))
```

```
## # A tibble: 2 x 3
##   hispanic     n   prop
##   <lgl>    <int>  <dbl>
## 1 FALSE     1853 0.949
## 2 TRUE        99 0.0507
```

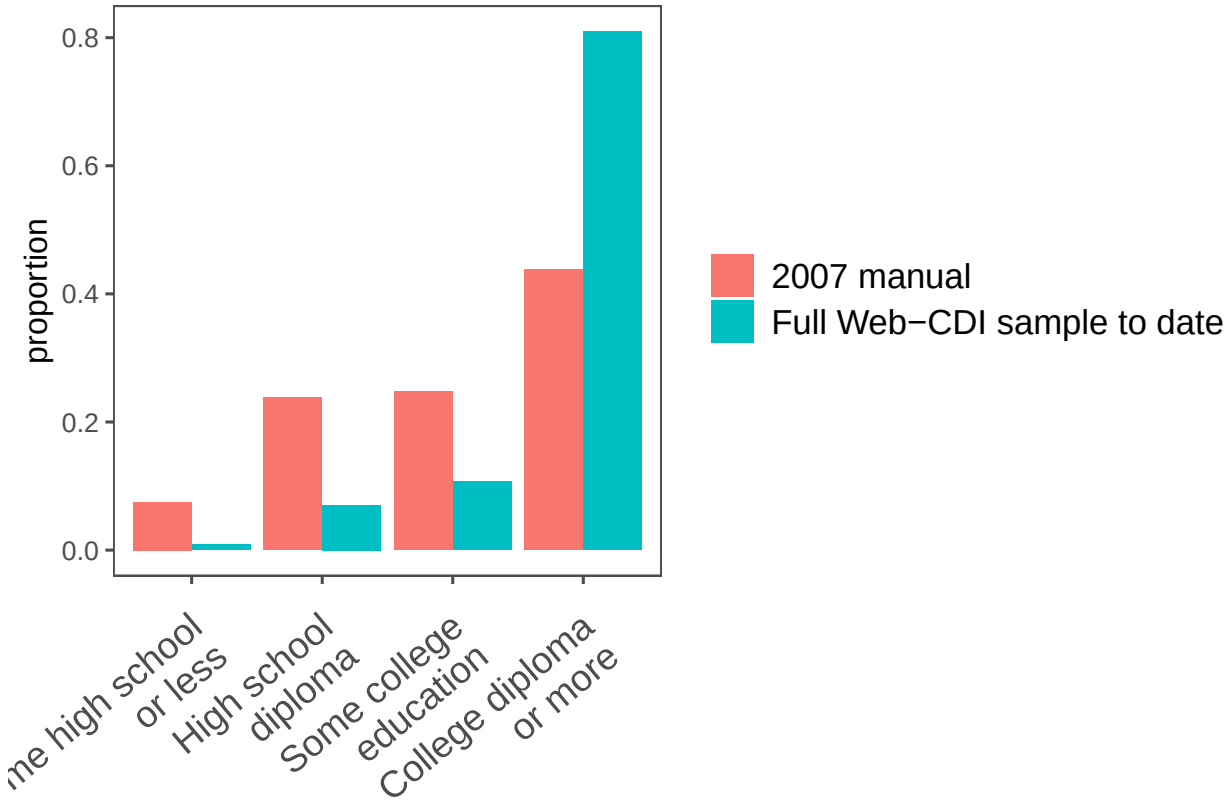
```
#Demographic analyses on the entire sample


demographics_df <-
  bind_rows(
    wg_filtered %>%
      select(
        study_name,
        subject_id,
        age,
        ethnicity,
        maternal_ed,
        words_produced = `Words Produced`
      ),
    ws_filtered %>%
      select(
        study_name,
        subject_id,
        age,
        ethnicity,
        maternal_ed,
        words_produced = `Total Produced`
      )
  )

ethnicity_plot_df <-
  demographics_df %>%
  getEthnicitySummary() %>%
  filter(!is.na(ethnicity)) %>%
  filter(ethnicity != "No ethnicity reported") %>%
  mutate(`Web-CDI sample` = prop.table(n)) %>%
  left_join(old_ethnicity_numbers, by = "ethnicity") %>%
```

```
  select(-n) %>%
  pivot_longer(
    cols = c(`Web-CDI sample`, `2007 manual`),
    names_to = "study",
    values_to = "proportion"
  )

ethnicity_plot <-
  ethnicity_plot_df %>%
  ggplot(aes(ethnicity, proportion, fill = study)) +
  geom_col(position = "dodge") +
  labs(
    y = "Proportion of\nrespondents"
  ) +
  theme_few() +
  theme(
    legend.title = element_blank(),
    axis.text = element_text(size = 14),
    axis.title = element_text(size = 13),
    legend.text = element_text(size = 13),
    axis.title.x = element_blank(),
    plot.title = element_text(size = 15),
    plot.caption = element_text(hjust = 0)
  )

ethnicity_plot
```




```
#Maternal ed analysis on the full sample

maternal_ed_plot_df <-
  demographics_df %>%
  count(maternal_ed) %>%
  mutate(`Full Web-CDI sample to date` = prop.table(n)) %>%
  left_join(old_momed_numbers, by = "maternal_ed") %>%
  select(-n) %>%
  pivot_longer(
    cols = c(`Full Web-CDI sample to date`, `2007 manual`),
    names_to = "study",
    values_to = "proportion"
  ) %>%
  mutate(
    maternal_ed = fct_relevel(
      maternal_ed,
      "Some high school or less",
      "High school diploma",
      "Some college education",
      "College diploma or more"
    )
  ) %>%
  filter(!is.na(maternal_ed))

x_axis_labs <- c(
  "Some high school\n or less",
```

```
  "High school\ndiploma",
  "Some college\neducation",
  "College diploma\nor more"
)

maternal_ed_plot <-
  maternal_ed_plot_df %>%
  filter(maternal_ed != "Not reported") %>%
  ggplot(aes(maternal_ed, proportion, fill = study)) +
  geom_col(position = "dodge") +
  theme_few() +
  theme(
    legend.title = element_blank(),
    axis.text.x = element_text(angle = 40, vjust = 0.9, hjust = 1, size = 13.5),
    axis.title.x = element_blank(),
    legend.text = element_text(size = 13),
    plot.caption = element_text(hjust = 0)
  ) +
  scale_x_discrete(labels = x_axis_labs)

maternal_ed_plot
```




```
#this table can be printed out
maternal_ed_table <-
  maternal_ed_plot_df %>%
```

```
  filter(maternal_ed != "Not reported") %>%
  pivot_wider(names_from = "study", values_from = "proportion") %>%
  mutate(
    `Current study proportions` = round(`Full Web-CDI sample to date`, digits = 4)
  ) %>%
  select(-`Full Web-CDI sample to date`)
```

`

```
#More fine grained analyses of exclusions. Copied and pasted from all_norming_analysis.Rmd.

n_total_wg <- nrow(all_wg_raw)
n_total_ws <- nrow(all_ws_raw)

excl_col_names <-
  c(
    "Exclusion",
    "WG exclusions",
    "% of full WG sample excluded",
    "WS exclusions",
    "% of full WS sample excluded"
  )


#First take away kids who have done the survey more than once.
wg_minus_repeats <-
  all_wg_raw %>%
  getCompletionInterval() %>%
  filter(repeat_num == "1")

wg_repeats_n <- n_total_wg - nrow(wg_minus_repeats)

ws_minus_repeats <-
  all_ws_raw %>%
  getCompletionInterval() %>%
  filter(repeat_num == "1")

ws_repeats_n <- n_total_ws - nrow(ws_minus_repeats)

repeat_admins <-
  c(
    "Not first administration",
    wg_repeats_n,
    percent(wg_repeats_n / n_total_wg),
    ws_repeats_n,
    percent(ws_repeats_n / n_total_ws)
  )

names(repeat_admins) <- excl_col_names

#Next take away kids born pre-term or with low birthweight.

wg_minus_premie <-
```

```
  wg_minus_repeats %>%
  filterBirthweight()

wg_premie_n <- nrow(wg_minus_repeats) - nrow(wg_minus_premie)

ws_minus_premie <-
  ws_minus_repeats %>%
  filterBirthweight()

ws_premie_n <- nrow(ws_minus_repeats) - nrow(ws_minus_premie)

premies <-
  c(
    "Premature or low birthweight",
    wg_premie_n,
    percent(wg_premie_n / n_total_wg),
    ws_premie_n,
    percent(ws_premie_n / n_total_ws)
  )

names(premies) <- excl_col_names

#Next take away kids with multilingual exposure

wg_minus_multiling <-
  wg_minus_premie %>%
  filterMultilingual()

wg_multiling_n <- nrow(wg_minus_premie) - nrow(wg_minus_multiling)

ws_minus_multiling <-
  ws_minus_premie %>%
  filterMultilingual()

ws_multiling_n <- nrow(ws_minus_premie) - nrow(ws_minus_multiling)

multiling <-
  c(
    "Multilingual exposure",
    wg_multiling_n,
    percent(wg_multiling_n / n_total_wg),
    ws_multiling_n,
    percent(ws_multiling_n / n_total_ws)
  )

names(multiling) <- excl_col_names

#Next exclude kids with problems of illness, vision, or hearing
wg_minus_health <-
  wg_minus_multiling %>%
  filterIllnesses() %>%
  filterVision() %>%
  filterHearing()
```

```r
wg_health_n <- nrow(wg_minus_multiling) - nrow(wg_minus_health)

ws_minus_health <-
  ws_minus_multiling %>%
  filterIllnesses() %>%
  filterVision() %>%
  filterHearing()

ws_health_n <- nrow(ws_minus_multiling) - nrow(ws_minus_health)

health <-
  c(
    "Illnesses/Vision/Hearing",
    wg_health_n,
    percent(wg_health_n / n_total_wg),
    ws_health_n,
    percent(ws_health_n / n_total_ws)
  )

names(health) <- excl_col_names

#Now filter out kids who are the wrong age
wg_minus_age <-
  wg_minus_health %>%
  filter_age_wg()

wg_age_n <- nrow(wg_minus_health) - nrow(wg_minus_age)

ws_minus_age <-
  ws_minus_health %>%
  filter_age_ws()

ws_age_n <- nrow(ws_minus_health) - nrow(ws_minus_age)

age <-
  c(
    "Out of age range",
    wg_age_n,
    percent(wg_age_n / n_total_wg),
    ws_age_n,
    percent(ws_age_n / n_total_ws)
  )

names(age) <- excl_col_names

#Now we need to get rid of people who did the survey too fast

wg_minus_fakes <-
  wg_minus_age %>%
  filter(completion_time >= min_completion_time)

wg_fake_n <- nrow(wg_minus_age) - nrow(wg_minus_fakes)
```

```r
ws_minus_fakes <-
  ws_minus_age %>%
  filter(completion_time >= min_completion_time)

ws_fake_n <- nrow(ws_minus_age) - nrow(ws_minus_fakes)

fakes <-
  c(
    "Completed survey too quickly",
    wg_fake_n,
    percent(wg_fake_n / n_total_wg),
    ws_fake_n,
    percent(ws_fake_n / n_total_ws)
  )

names(fakes) <- excl_col_names

#calculate total amount of WG exclusions
total_wg_exclusions <-
  wg_repeats_n +
  wg_premie_n +
  wg_multiling_n +
  wg_health_n +
  wg_age_n +
  wg_fake_n

#calculate total amount of WS exclusions
total_ws_exclusions <-
  ws_repeats_n +
  ws_premie_n +
  ws_multiling_n +
  ws_health_n +
  ws_age_n +
  ws_fake_n

#make a row in the table for this
totals <-
  c(
    "Total exclusions",
    total_wg_exclusions,
    percent(total_wg_exclusions / n_total_wg),
    total_ws_exclusions,
    percent(total_ws_exclusions / n_total_ws)
  )

names(totals) <- excl_col_names

#now make the table
exclusion_tbl <-
  bind_rows(repeat_admins, premies, multiling, health, age, fakes, totals)

knitr::kable(exclusion_tbl)
```

Exclusion	WG exclusions	% of full WG sample excluded	WS exclusions	% of full WS sample excluded
Not first administration	163	6%	444	12%
Premature or low birthweight	37	1%	67	2%
Multilingual exposure	449	16%	492	14%
Illnesses/Vision/Hearing	191	7%	203	6%
Out of age range	88	3%	200	6%
Completed survey too quickly	363	13%	236	7%
Total exclusions	1291	45%	1642	46%