NF1 Knowledge Analysis

Marshall McQuillen

8/19/2019

Contents

Overview **1**

Statistical Analysis **2**
- Study Aims . . . 2
- Data Overview & Cleaning . . . 2
- Approach 1 . . . 2
- Approach 2 . . . 8
- Cronbach's Alpha . . . 10
- Statistical Theory: Approach 1 vs. Approach 2 . . . 11

Data Visualizations **13**

References **26**

Overview

The paper that this analysis was created for was submitted for publication in "The Journal Of Genetic Counseling." As the journal name implies, it's focus is on genetic counseling and not on the field of statistics, and therefore some of the nuances of the statistics had to be left out. This report will take a closer look at those nuances. The paper that was submitted is titled, "Assessment of the Impact of a Positive Family History and Genetic Counseling on Parental Knowledge of Neurofibromatosis Type 1 (NF1)."

Statistical Analysis

Study Aims

The study had two hypotheses, namely:

1. Parental knowledge of NF1 is influenced by having a personal diagnosis of NF1.
2. Parental knowledge of NF1 can be increased by exposure to genetic counseling.

Data Overview & Cleaning

Prior to performing any statistical tests, some minor data cleaning was needed. The below lists all alterations made to the data for the purposes of this and the submitted paper.

There were a total of 303 responses to the questionare:

- 11 respondents were "Not sure" if they had NF1. These respondents were dropped from the data set since both study aims are contingent upon knowing this information.
- One respondent skipped the question, "Does your child's other parent have NF1?" This respondent was kept since they answered the question, "Do you have NF1?" and their answer to "Does your child's other parent have NF1?" was imputed as "No."
- One respondent skipped the question, "Doe you have NF1?" This respondent was dropped from the data set since both study aims are contingent upon knowing this information.
- 14 respondents do not have NF1 and are "Not sure," if the childs other parent has NF1. These respondents were dropped from the data set since both aims of the thesis are dependent on this information.
- Three additional respondents were dropped from the study due to leaving other questions that will be discussed blank.

After the above adjustments, the final data set contains 274 respondents, with an age ranging from 25 to 82 and a mean age of 42.1135531. The standard deviation of the ages is 9.9638095.

Approach 1

Testing Methodology

In order to assess the first hypothesis of the study, to determine whether parental NF1 knowledge is influenced by the parent having a personal diagnosis of NF1, respondents were asked whether they or their affected child's other parent has NF1. This split the respondents into three groups:

- The **Unaffected** group - Respondents who do not have NF1 and the child's other parent also does not have NF1.
- The **Other Parent Affected** group - Respondents who do not have NF1 and the child's other parent does have NF1.
- The **Affected** group - Respondents who do have NF1 and the child's other parent does not have NF1.

Originally, since we (myself and the PI of the study) knew we wanted to compare each of the three groups to the other two, I decided to forgo an initial ANOVA and go straight to the pairwise comparisons. Therefore, the three tests, and their accompanying hypotheses are as follows:

1. Test 1 hypothesis: The "Unaffected" group and the "Other Parent Affected" group do not have the same mean NF1 knowledge score, as measured by the knowledge questionaire.
   - $H_0: \mu_{unaffected} = \mu_{other\ parent\ affected}$
   - $H_a: \mu_{unaffected} \neq \mu_{other\ parent\ affected}$
2. Test 2 hypothesis: The "Unaffected" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is greater than that of the "Affected" group.
   - $H_0: \mu_{unaffected} \leq \mu_{affected}$
   - $H_a: \mu_{unaffected} > \mu_{affected}$
3. Test 3 hypothesis: The "Other Parent Affected" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is greater than that of the "Affected" group.
   - $H_0: \mu_{other\ parent\ affected} \leq \mu_{affected}$
   - $H_a: \mu_{other\ parent\ affected} > \mu_{affected}$

Although the variance of test scores are very close to each other, the sample sizes are not (see *Table 1. Study Aim 1: Sample Statistics*), leading to the choice of Welch's t-test as the statistical test to be used since it doesn't assume equal variances nor equal sample sizes.

The second hypothesis of the study, to determine whether parental knowledge of NF1 can be increased by exposure to genetic counseling, was also tested using Welch's t-test for the same reasons as mentioned above (see *Table 2. Study Aim 2: Sample Statistics*). The hypothesis for this test are below.

4. Test 4 hypothesis: The mean NF1 knowledge score, as measured by the knowledge questionaire, of those who have had exposure to genetic counseling is greater than that of those who have not had exposure to genetic counseling.
   - $H_0: \mu_{seen\ GC} \leq \mu_{not\ seen\ GC}$
   - $H_a: \mu_{seen\ GC} > \mu_{not\ seen\ GC}$

In order to control the Family-Wise Type I error rate, the original significance level ($\alpha = 0.05$) was adjusted using the Bonferroni Correction, reducing the level of statistical significance to $\alpha_{adj} = 0.0125$.

Table 1: Study Aim 1: Sample Statistics

	Mean	Variance	N
Unaffected	0.8276901	0.0084524	193
Other Parent Affected	0.8436247	0.0056361	29
Affected	0.7768336	0.0075135	52

Table 2: Study Aim 2: Sample Statistics

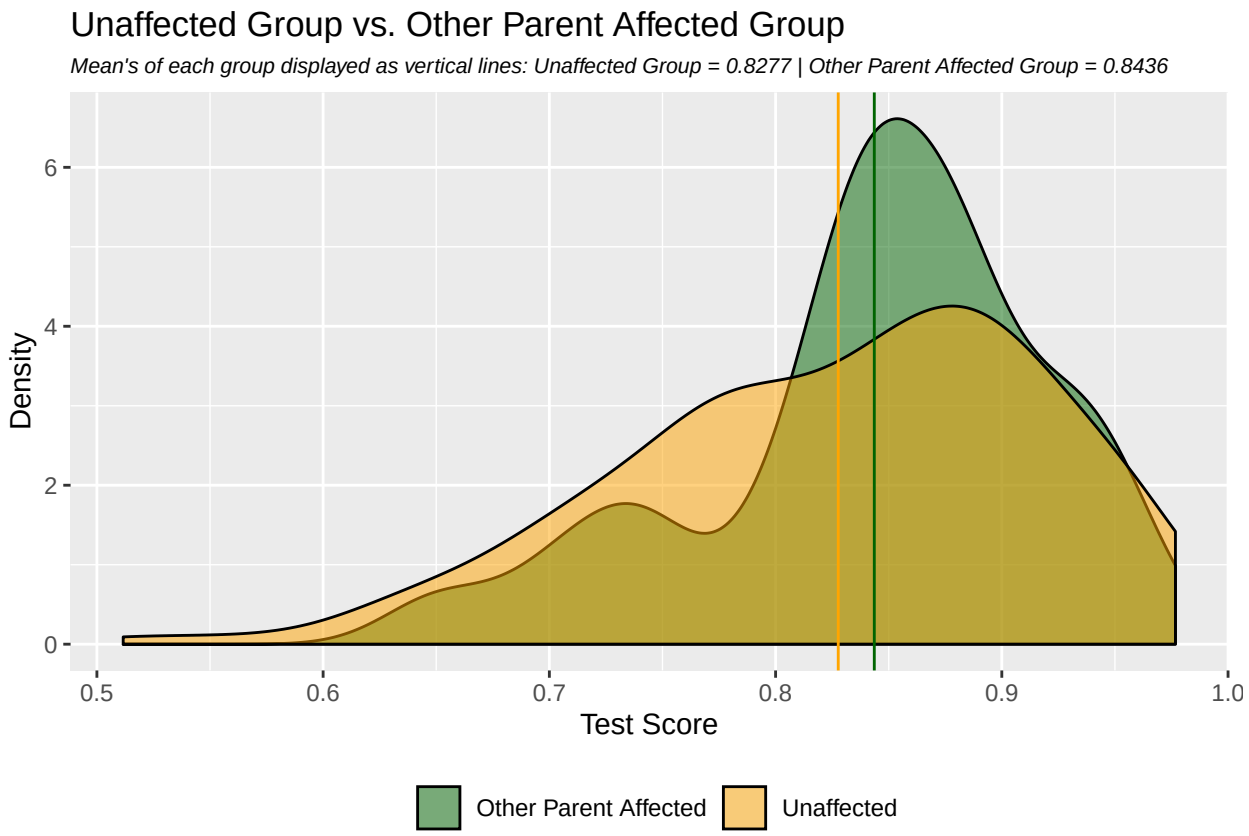
	Mean	Variance	N
Has seen a Genetic Counselor	0.8315752	0.0075986	194
Has not seen a Genetic Counselor	0.7909884	0.0092215	80

This brings us to the tests for the first study aim.

Study Aim 1 - Parental knowledge of NF1 is influenced by having a personal diagnosis of NF1.

Test 1

- H_0 : The "Unaffected" group has the same mean NF1 knowledge score, as measured by the knowledge questionaire, as the "Other Parent Affected" group.
- H_a : The "Unaffected" group does not have the same mean NF1 knowledge score, as measured by the knowledge questionaire, as the "Other Parent Affected" group.

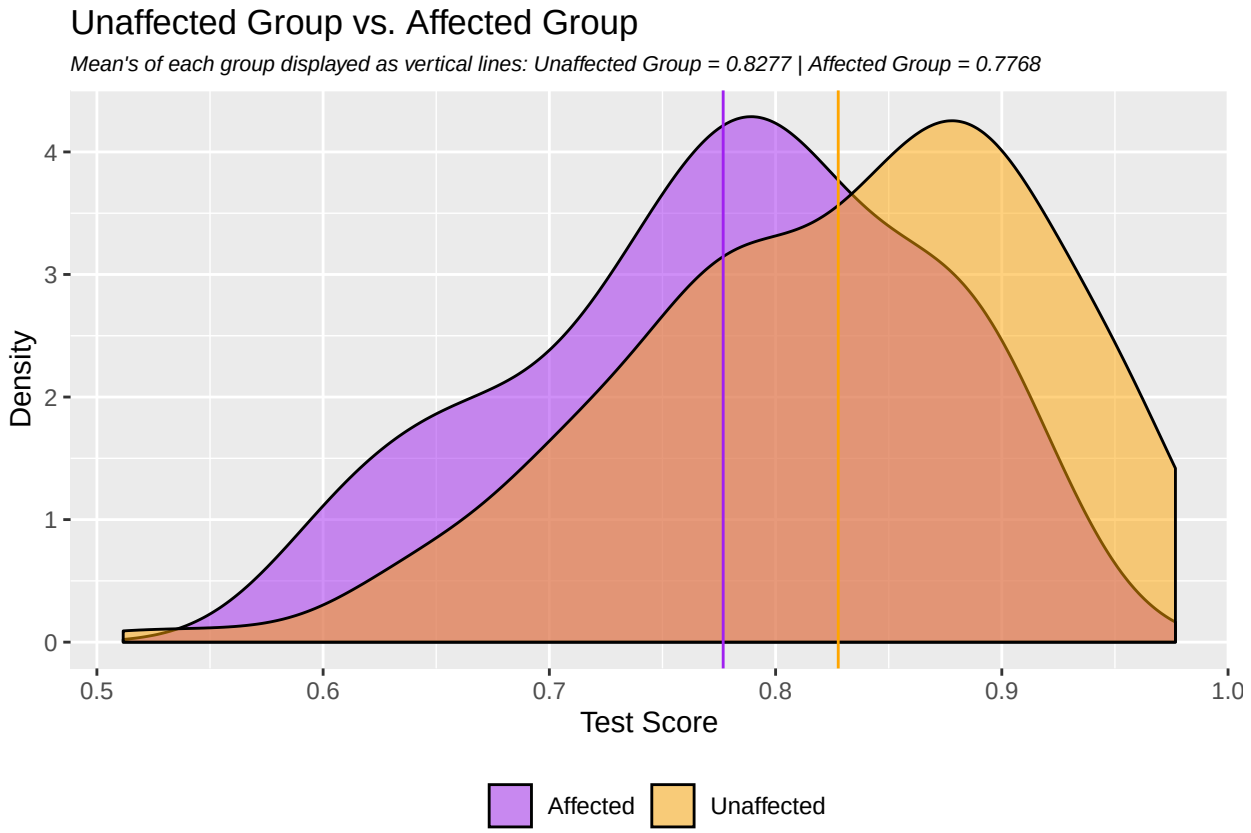



```
## [1] "Test 1 p value = 0.307749444337635"
```

Welch's two-tailed t-Test returns the above p value, with which we fail to reject the null hypothesis.

Test 2

- H_0 : The "Unaffected" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is less than or equal to that of the "Other Parent Affected" group.
- H_a : The "Unaffected" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is greater than that of the "Affected" group.

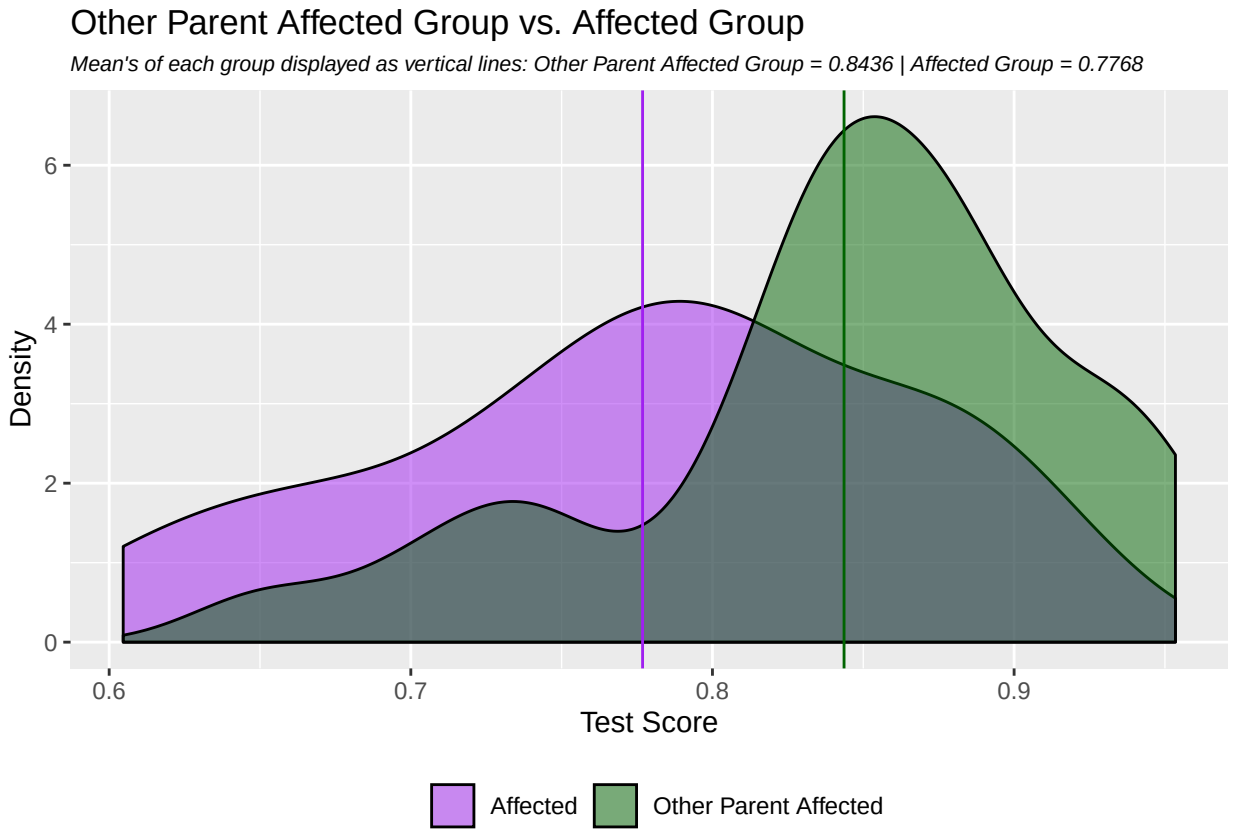



```
## [1] "Test 2 p value = 0.000187363981785434"
```

Welch's one-sided t-test returns the above p value. Being well below the Bonferroni corrected significance level of 0.0125, H_0 is rejected in favor H_a.

Test 3

- H_0 : The "Other Parent Affected" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is less than or equal to that of the "Affected" group.
- H_a : The "Other Parent Affected" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is greater than that of the "Affected" group.



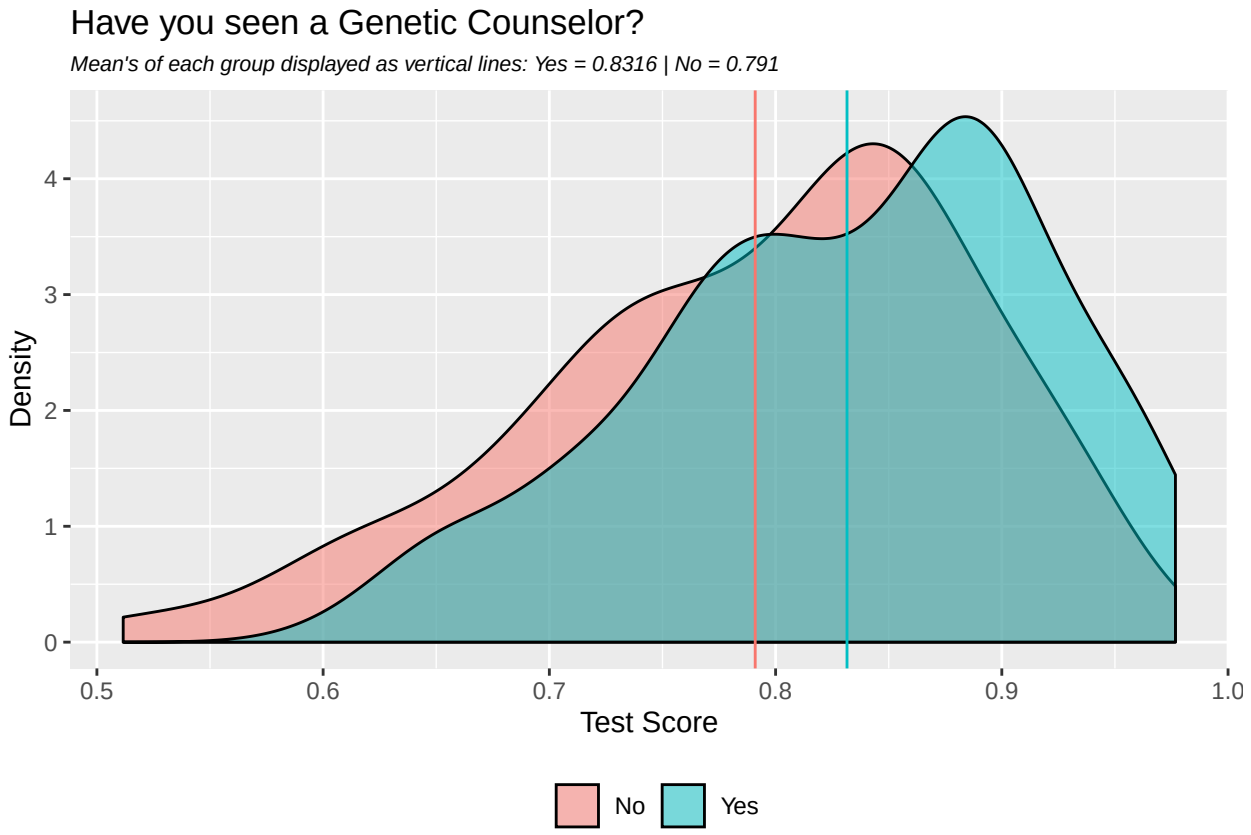

```
## [1] "Test 3 p value = 0.000280154187709805"
```

Welch's one-sided t-Test returns the above p value. Being well below the Bonferroni corrected significance level of 0.0125, H_0 is rejected in favor H_a.

Study Aim 2 - Parental knowledge of NF1 can be increased by exposure to genetic counseling.

Test 4

- H_0 : The "Seen GC" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is less than or equal to that of the "Not Seen GC" group.
- H_a : The "Seen GC" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is greater than that of the "Not Seen GC" group.




```
## [1] "Test 4 p value = 0.000691029479878431"
```

Welch's one-sided t-Test returns the above p value. Being well below the Bonferroni corrected significance level of 0.0125, H_0 is rejected in favor H_a.

Approach 2

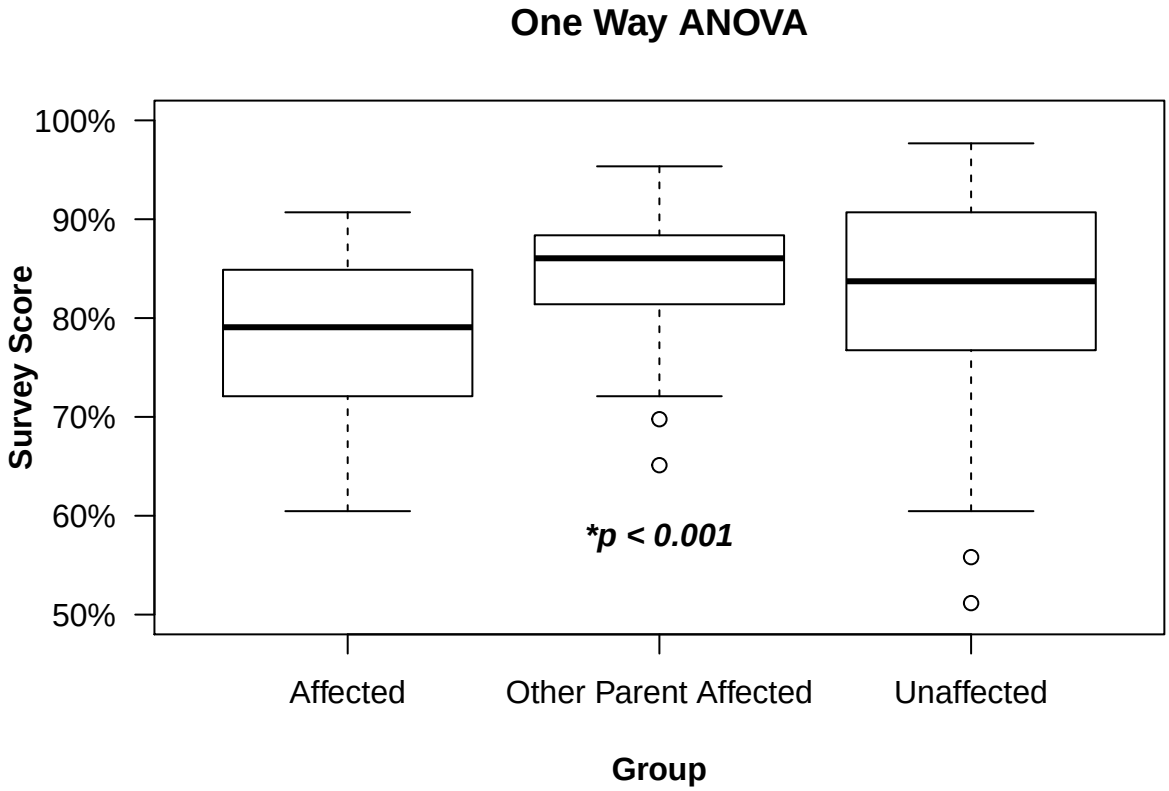
Upon initial submission of the paper, the editor of the journal noted that the "statistical norms" of studies in the Journal of Genetic Counseling (JOGC) are to use an ANOVA followed by post-hoc pairwise comparisons should the ANOVA prove significant. In addition, should the paper be published in black and white, the above images wouldn't be as clear as they could be. With these suggestions in mind, the analysis and images were redone to fit the guidelines of the journal.

Study Aim 1 - Parental knowledge of NF1 is influenced by having a personal diagnosis of NF1.

ANOVA

Since there are three groups that need to be tested ("Affected", "Other Parent Affected" and "Unaffected"), an ANOVA test will determine whether there is evidence that at least one group mean is different from the others.

- H_0: There is no difference in mean NF1 knowledge score, as measured by the knowledge questionaire, between the three groups ("Affected", "Other Parent Affected" and "Unaffected").
- H_a: At least one group has a different mean NF1 knowledge score, as measured by the knowledge questionaire, than the other groups.




```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## group_id      2 0.1245 0.06224   7.794 0.000511 ***
## Residuals   271 2.1639 0.00798
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

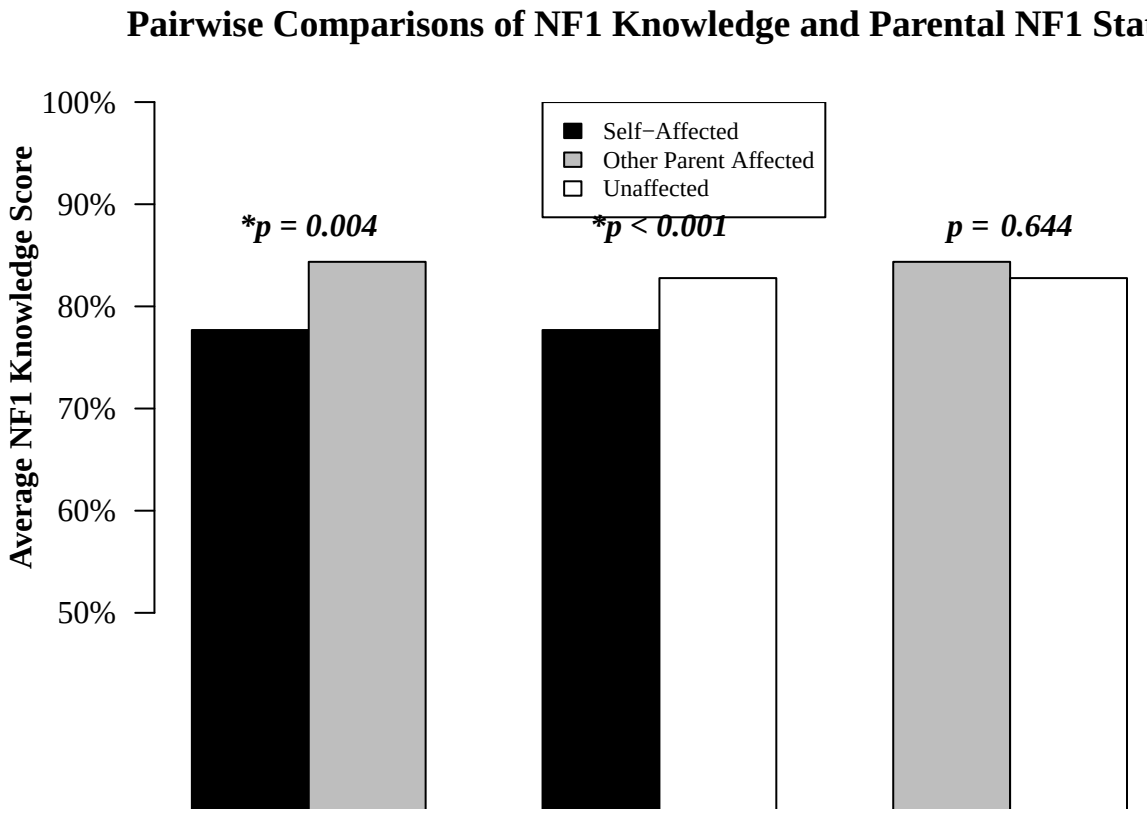
The one way ANOVA, $F(2, 271) = 7.794$, $p = 0.000511$, shows that there is ample evidence for the alternate hypothesis, which is accepted and followed up with the pairwise comparisons.

Tukey's HSD

Following the ANOVA result with Tukey's Honestly Significant Difference (HSD) test, it is shown that the "Affected" group has a significantly lower NF1 knowledge score, as measure by the knowledge questionaire, than the other two groups. A significant difference between the "Unaffected" and "Other Parent Affected" groups was not found.

Table 3: Tukey HSD Results

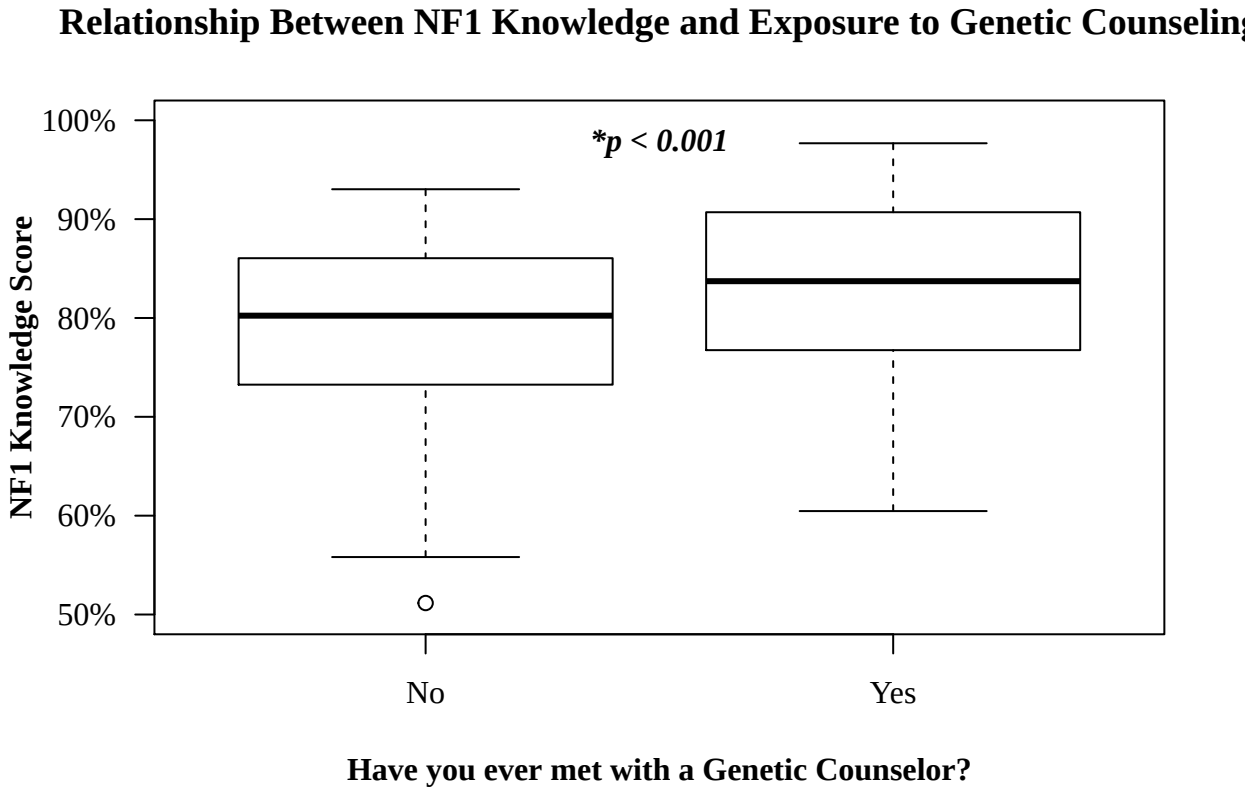
	(Adj.) P value
Other Parent Affected-Affected	0.0040266
Unaffected-Affected	0.0009406
Unaffected-Other Parent Affected	0.6436802



Study Aim 2 - Parental knowledge of NF1 can be increased by exposure to genetic counseling.

The testing methodology for the second study aim remained the same for both approaches, however the plot was changed to display the findings in a more traditional manner (hypotheses reproduced below)

- H_0 : The "Seen GC" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is less than or equal to that of the "Not Seen GC" group.
- H_a : The "Seen GC" group has a mean NF1 knowledge score, as measured by the knowledge questionaire, that is greater than that of the "Not Seen GC" group.



Cronbach's Alpha

Cronbach's alpha is a statistic used to measure the reliability of a set of test items, or more specifically, how well a set of test items (questions) measures the underlying concept which the test is attempting to assess. In the context of this thesis, Cronbach's Alpha can be interpreted as a numeric summary of how well the knowledge based questions assess a respondents true knowledge level of NF1.

Using 43 quiz questions returned a Cronbach's Alpha value of 0.7010484.

Statistical Theory: Approach 1 vs. Approach 2

The fact that both the above approaches produced the same result is a testament to the significance of the findings, however I believe it is worth diving into the theory of the two approaches, and in particular arguing that a one way ANOVA F-test was unnecessary in this case.

The null hypothesis of a one way ANOVA F-test is that the means of all groups tested are the same. That is,

$$H_o : \mu_1 = \mu_2 = \mu_3 \cdots \mu_k$$

Therefore, the alternate hypothesis of one way ANOVA is that *at least two* group means are unequal. That is,

$$H_a : at\ least\ one\ pair\ \mu_j \neq \mu_{j'}$$

Should a one way ANOVA F-test prove significant and the alternate hypothesis be accepted, the natural next question becomes, "Which pair(s) have different means?"

This is where post hoc comparisons come into play, such as the Bonferroni Correction, Fisher's Least Significant Difference (LSD) test and Tukey's Honestly Significant Difference (HSD) test (used in Approach 2), to name a few. While diving into the details of each of these methods is outside the scope of this paper and deserves focused research, their general purpose is to control the family-wise error rate (FWER), sometimes referred to as the experimentwise-alpha.

When one decides to correct for multiple comparisons, defining the bounds of a "family" of statistical tests has a large impact on the whether a particular test is deemed "significant," and, "there is no firm rule on this."[1] Some texts suggests that a "family" of tests are all tests, "conducted on the same data."[2] However, if this is the definition of "family" that is used, than post hoc tests inherently do *not* control the FWER by definition. Post hoc (latin: *After this*) tests, as the name implies, are performed after an initial test for significance is conducted (ANOVA in this and most cases) and ensure that the probability of a false positive among *any of the post hoc tests* does not exceed a certain threshold. Since there isn't a post hoc test that retroactively adjusts the significance level needed for the omnibus ANOVA to be deemed significant, post hoc procedures can only be said to control the FWER in so far as the "family" is defined as the pairwise post hoc tests. With this in mind, the ANOVA followed by post hoc tests method, the method used in Approach 2, has:

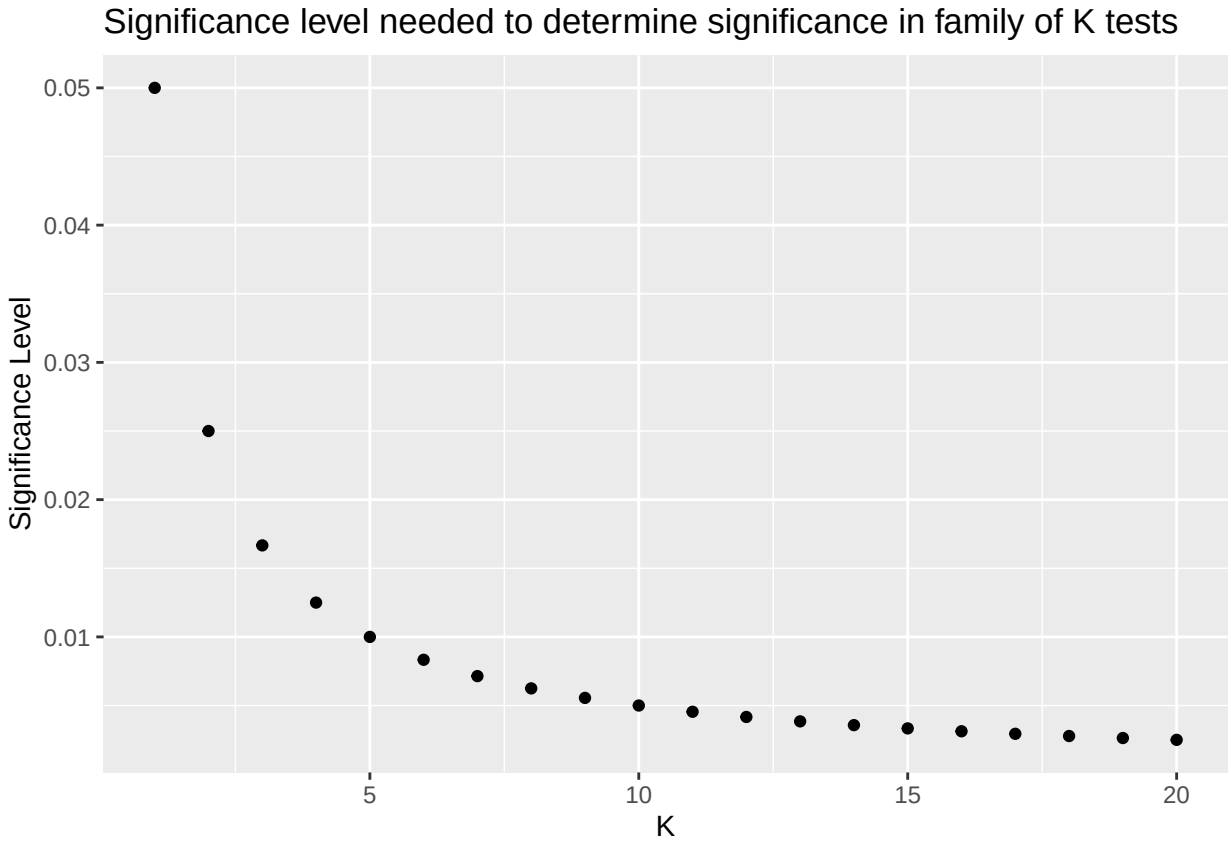
$$\begin{aligned} FWER &= probability\ of\ at\ least\ one\ Type\ I\ error\ in\ a\ family\ of\ hypothesis\ tests && (1)\\ &= P(H_a\ Anova\ falsely\ accepted) + P(H_a\ post\ hoc\ family\ falsely\ accepted) && (2)\\ &= 0.05 + 0.05 && (3)\\ &= 0.1 && (4) \end{aligned}$$

…if a "family" is defined as all tests that are conducted on the same sample.

In addition, in Approach 2, the adjustment to the significance level isn't extended to the test for the second study aim. Therefore, if included in the above calculation, the $FWER = 0.15$ after adding on the additional 0.05 for the test that looks at the second study aim.

Conversely, since we (myself and the PI) knew at the outset that we wanted to compare the means of each of the three groups, ("Affected", "Other Parent Affected" and "Unaffected"), in addition to the "Met GC" vs. "Not Met GC" groups, Approach 1 seemed more reasonable. By defining a "family" of tests as all those that are conducted on the same sample, applying the Bonferroni Correction to all 4 tests gives each test equal weight and does in fact control the FWER at 0.05. Although the Bonferroni Correction is considered

a convservative approach, this is outside the scope of this paper (again, deserves and has many of it's own papers), the significant levels only become overly conservative when $k > 4$.



This last section barely scratched the surface of a few questions in statistical theory, however, as I said, this would have been out of place in the JOGC.

Data Visualizations

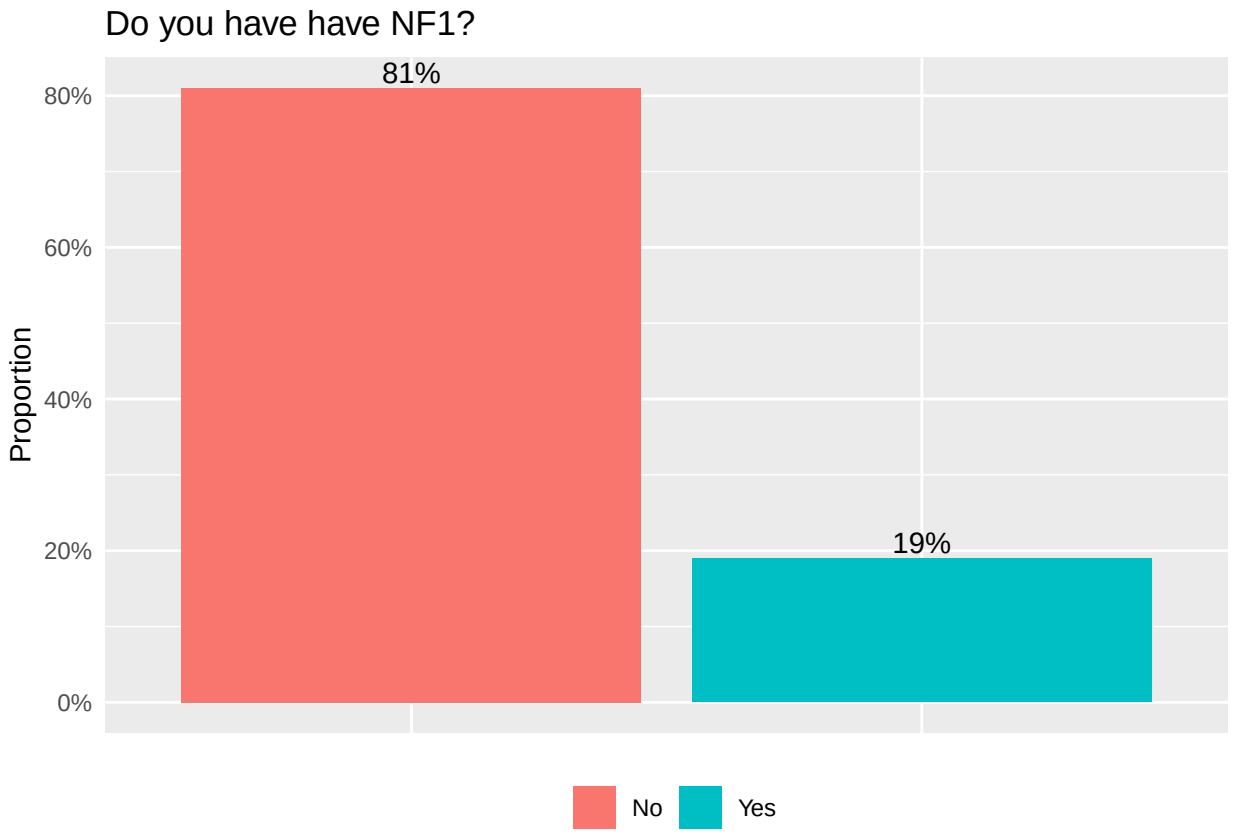


Table 4: Do you have NF1?

Response	Count
No	222
Yes	52

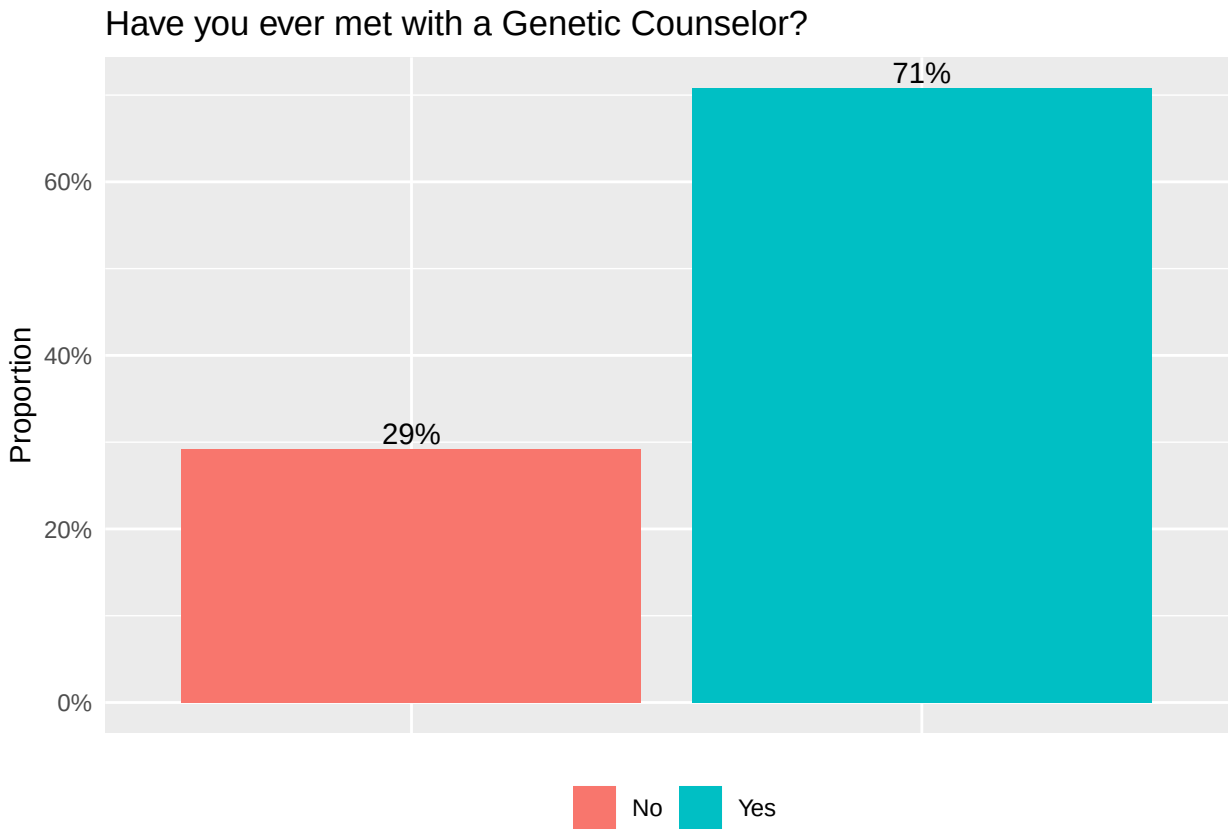


Table 5: Have you ever met with a Genetic Counselor?

Response	Count
No	80
Yes	194

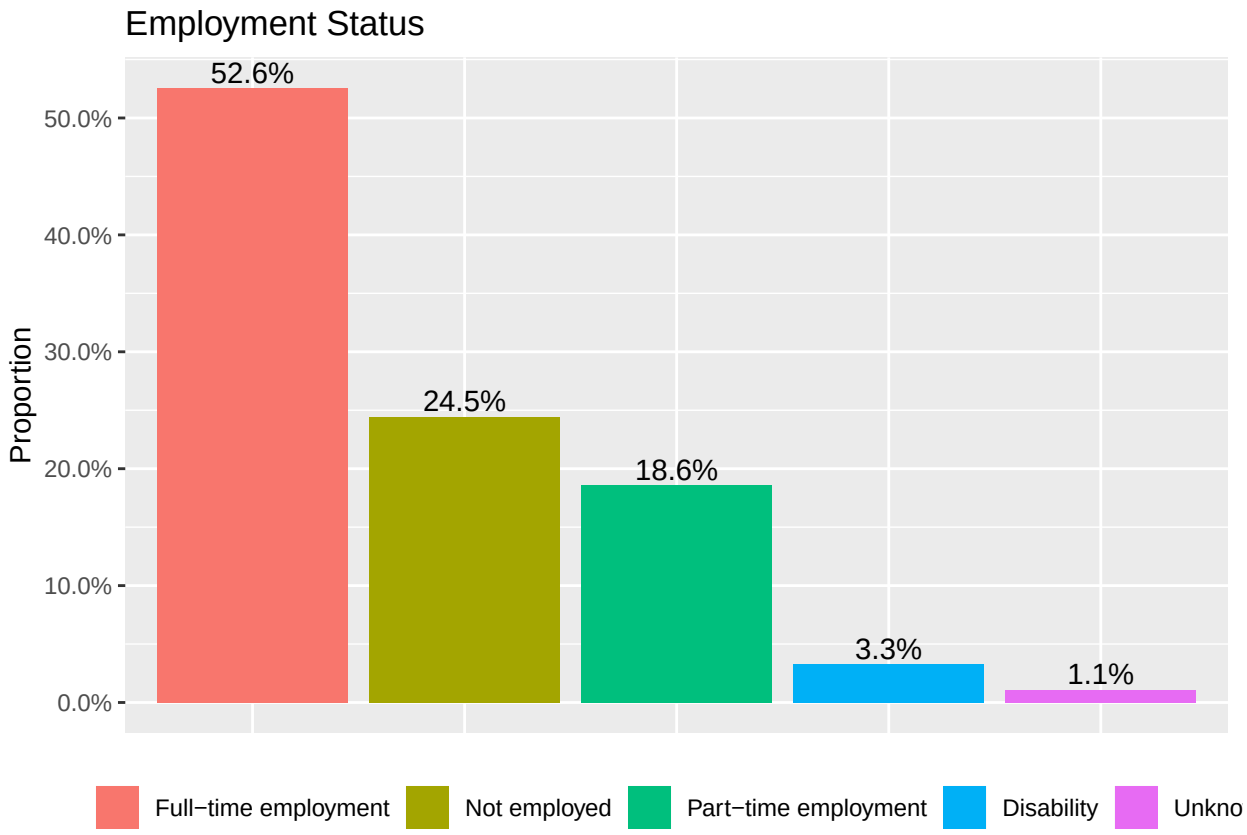


Table 6: What is your employment status?

Response	Count
Disability	9
Full-time employment	144
Not employed	67
Part-time employment	51
Unknown	3

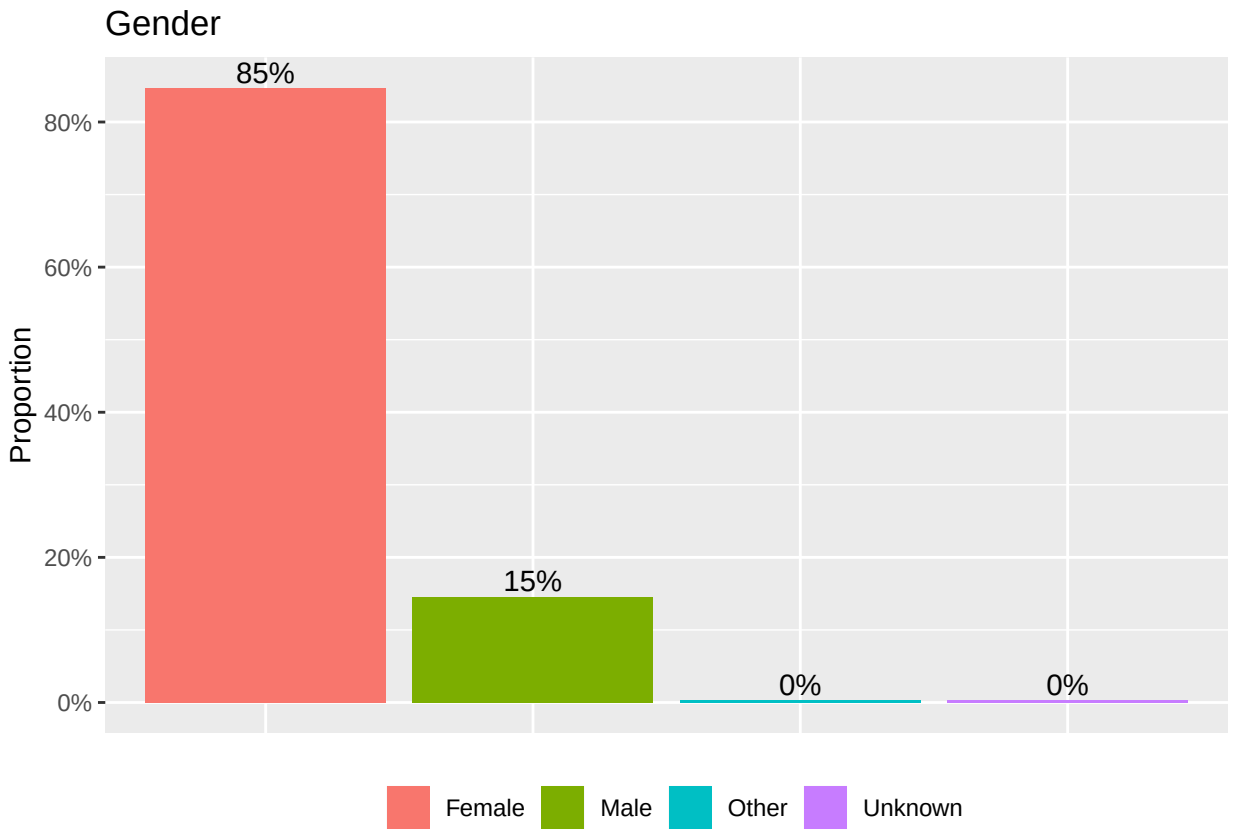


Table 7: Gender

Response	Count
Female	232
Male	40
Other	1
Unknown	1

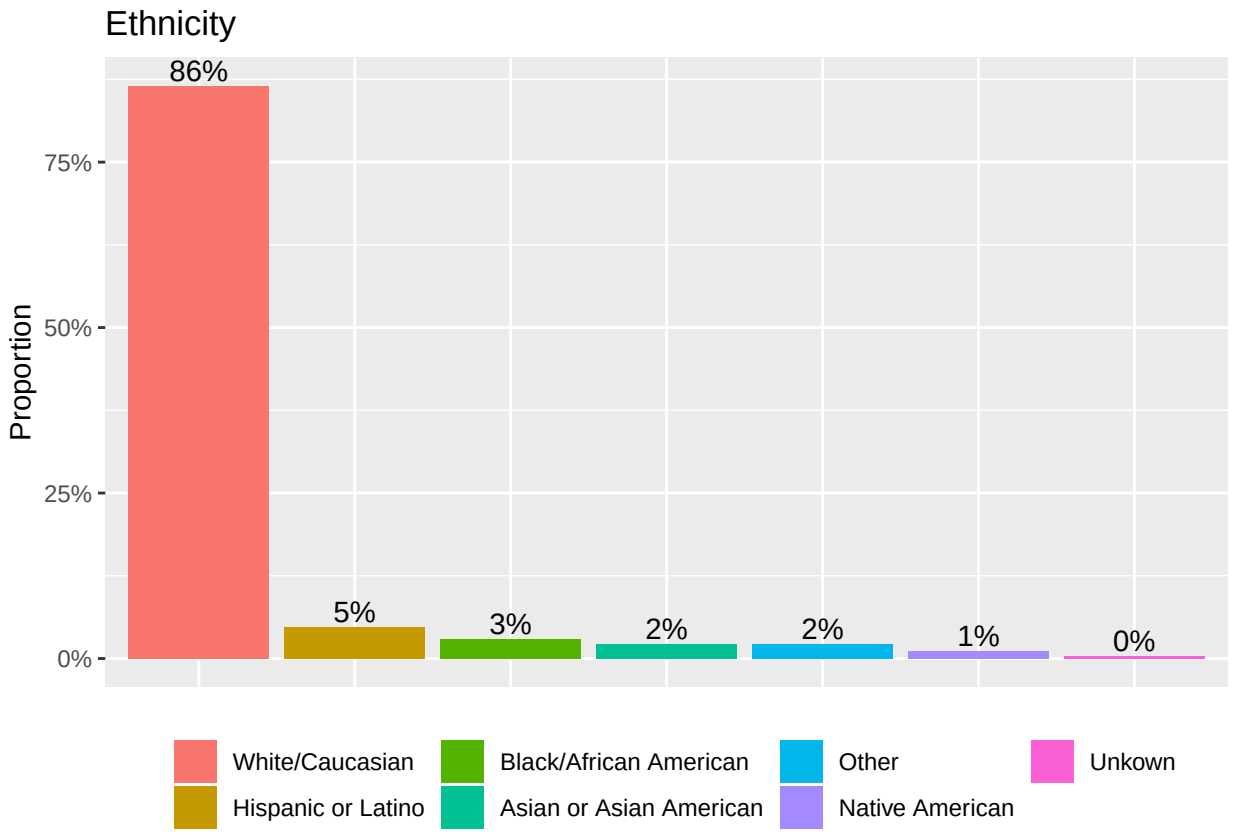


Table 8: Ethnicity

Response	Count
Asian or Asian American	6
Black/African American	8
Hispanic or Latino	13
Native American	3
Other	6
Unkown	1
White/Caucasian	237

One person did not answer the question regarding their educational background, and therefore was excluded from the below plot.

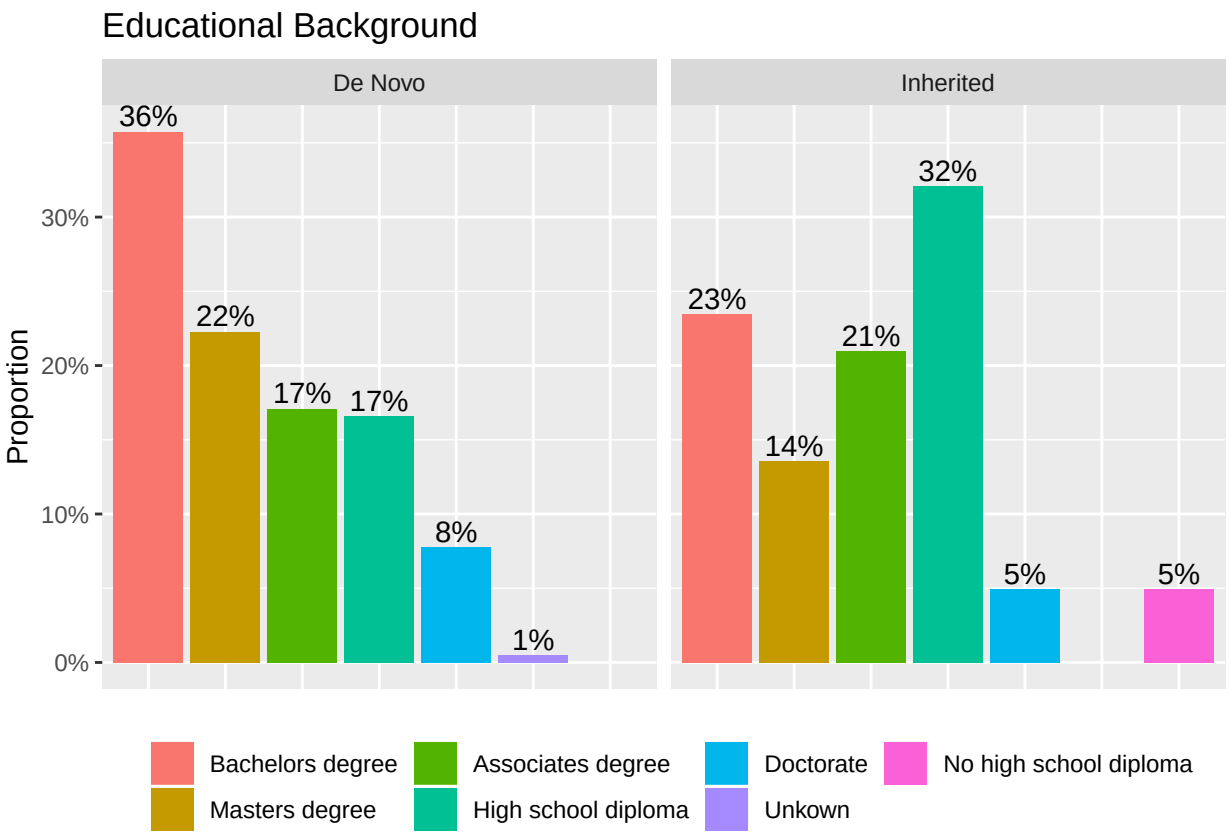
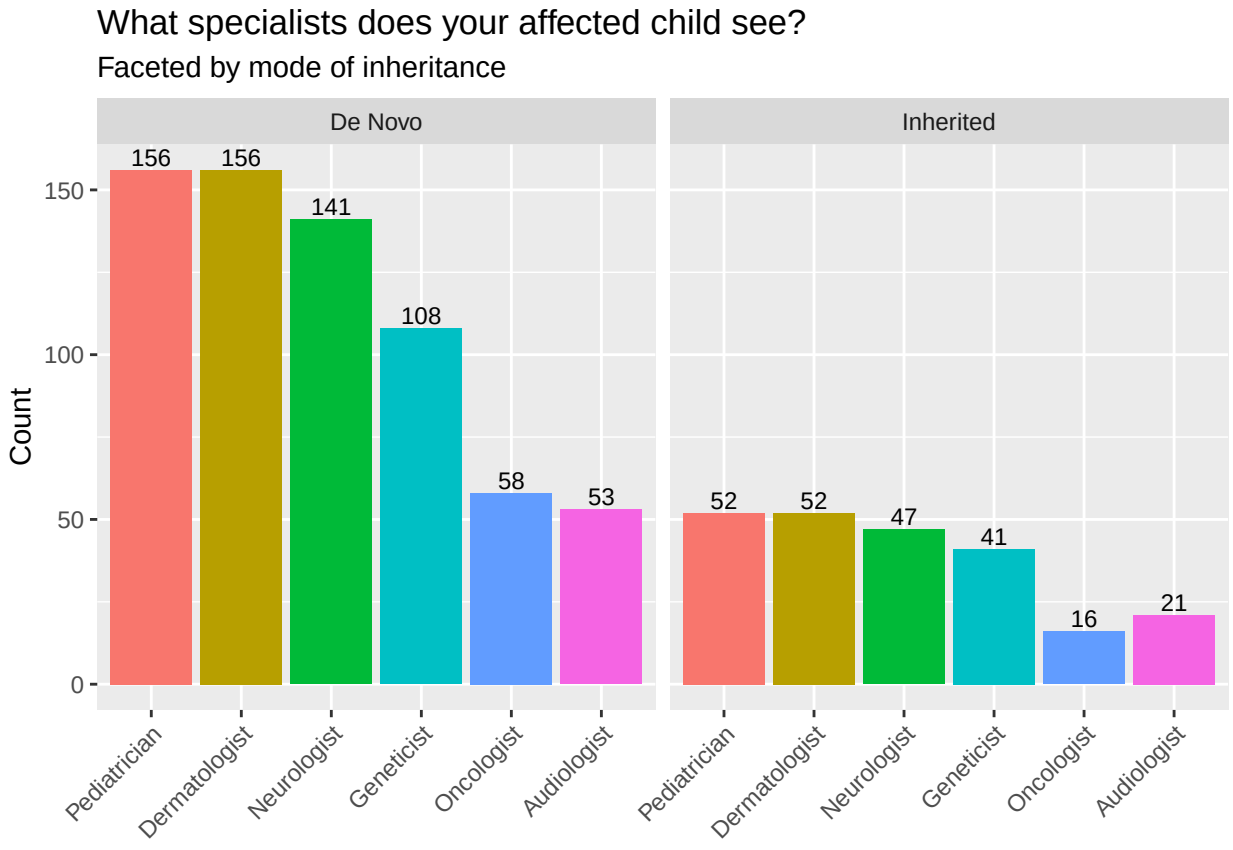


Table 9: Educational Background

Response	Count
Associates degree	50
Bachelors degree	88
Doctorate	19
High school diploma	58
Masters degree	54
No high school diploma	4

It is important to note that the answers to the below question, "What specialists does your affected child see?" are *not* mutually exclusive - children who see a Neurologist may have seen an Oncologist and a Geneticist as well.

```
## Warning in melt(display.df, id.vars = "mode_of_inheritance"): The melt generic
## in data.table has been passed a tbl_df and will attempt to redirect to the
## relevant reshape2 method; please note that reshape2 is deprecated, and this
## redirection is now deprecated as well. To continue using melt methods from
## reshape2 while both libraries are attached, e.g. melt.list, you can prepend the
## namespace like reshape2::melt(display.df). In the next version, this warning
## will become an error.
```



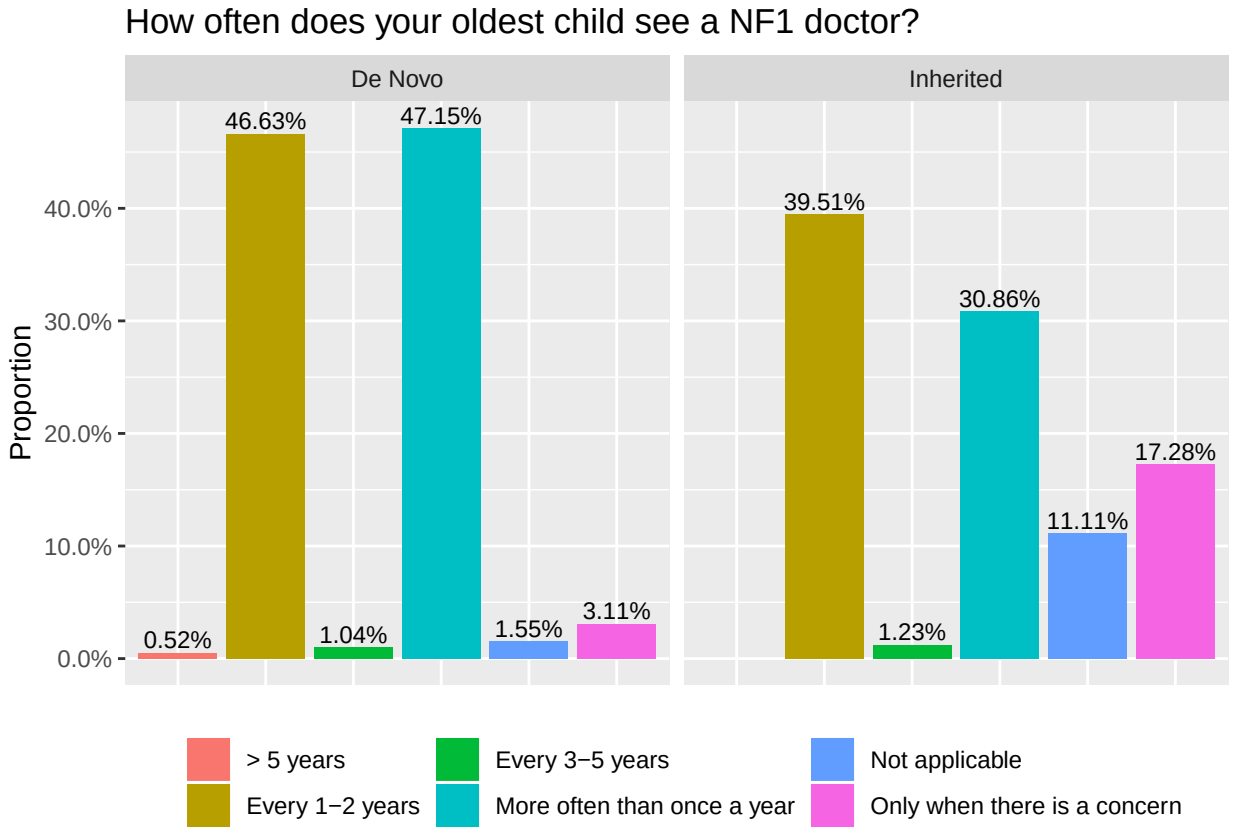
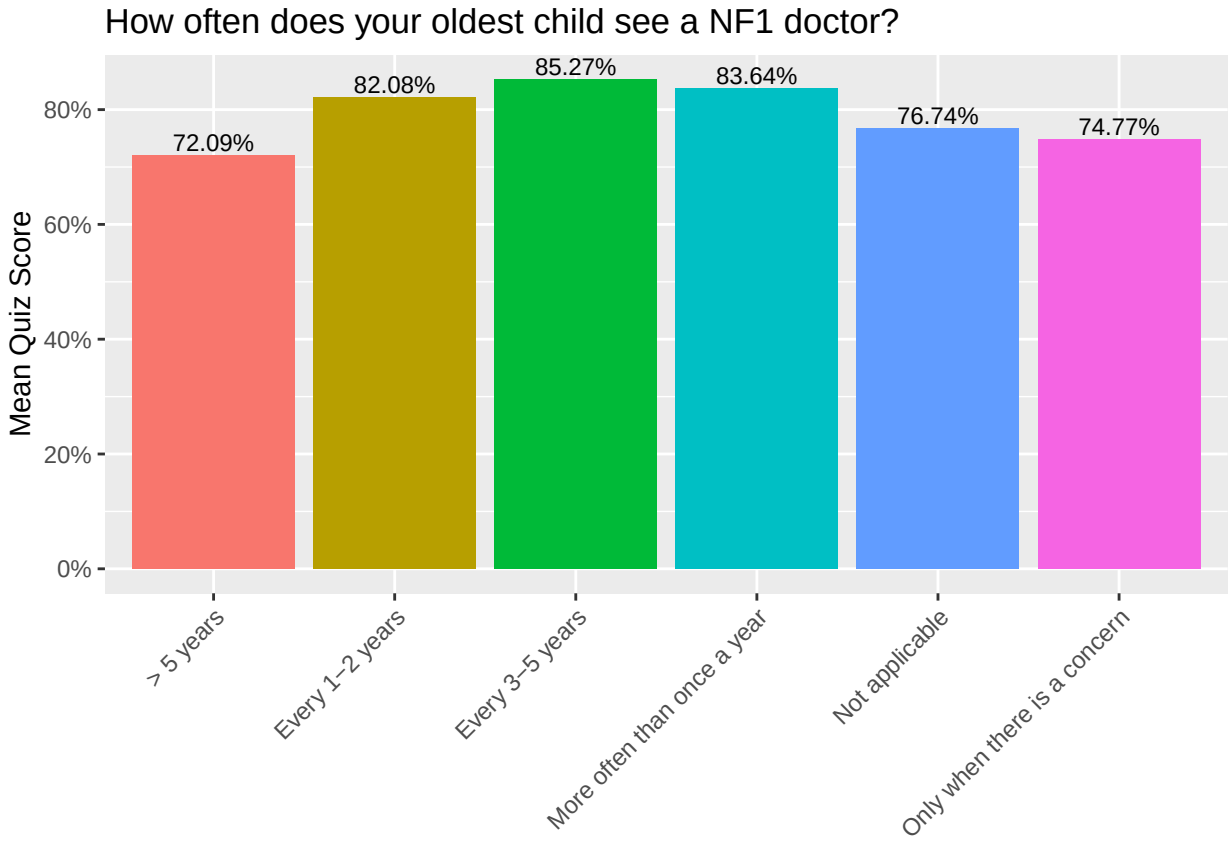
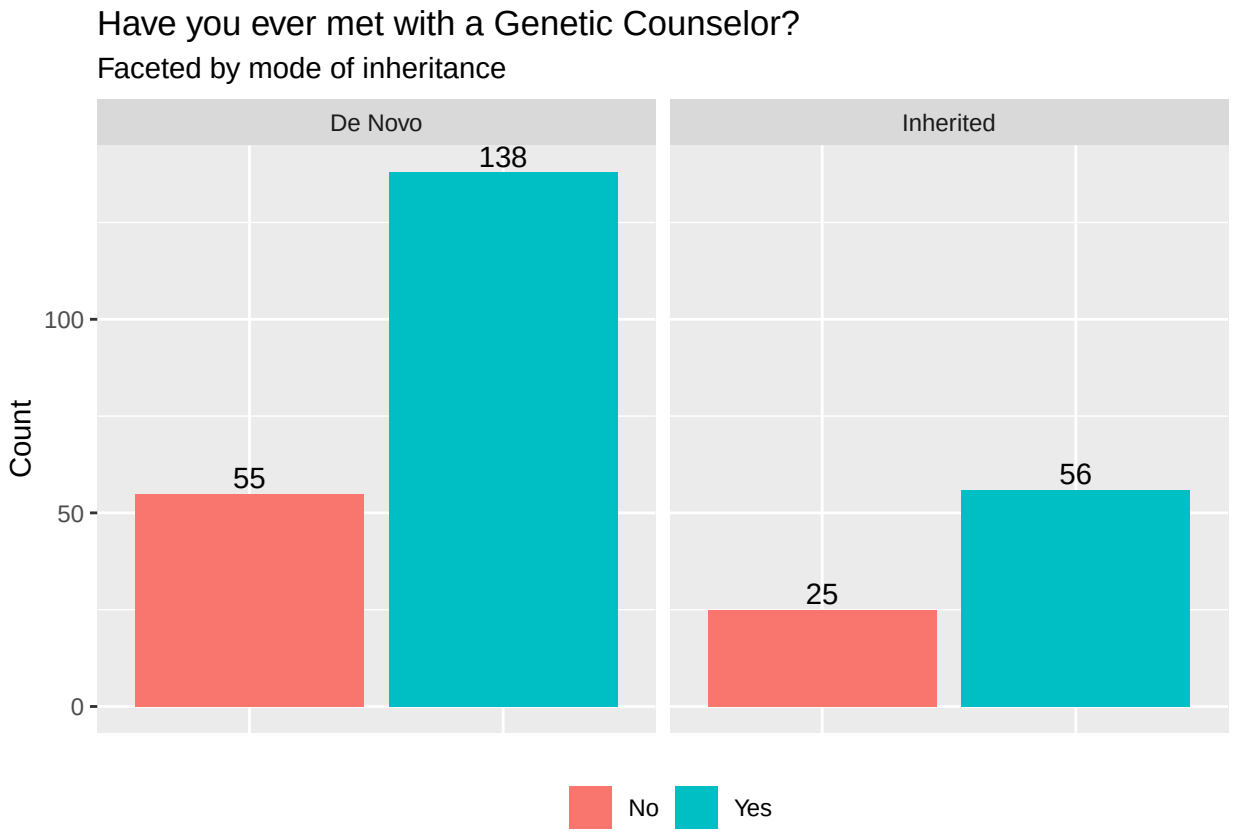


Table 10: How often does your oldest child see a NF1 doctor?

	De Novo	Inherited
> 5 years	1	0
Every 1-2 years	90	32
Every 3-5 years	2	1
More often than once a year	91	25
Not applicable	3	9
Only when there is a concern	6	14

```
## Warning in melt(display.df, id.vars =
## "how_often_does_your_.oldest._child_see_an_nf_doctor."): The melt generic in
## data.table has been passed a tbl_df and will attempt to redirect to the relevant
## reshape2 method; please note that reshape2 is deprecated, and this redirection
## is now deprecated as well. To continue using melt methods from reshape2 while
## both libraries are attached, e.g. melt.list, you can prepend the namespace like
## reshape2::melt(display.df). In the next version, this warning will become an
## error.
```




Have you ever met with a Genetic Counselor?
Faceted by mode of inheritance
De Novo
Inherited
Count
100
50
0
55
138
25
56
No
Yes

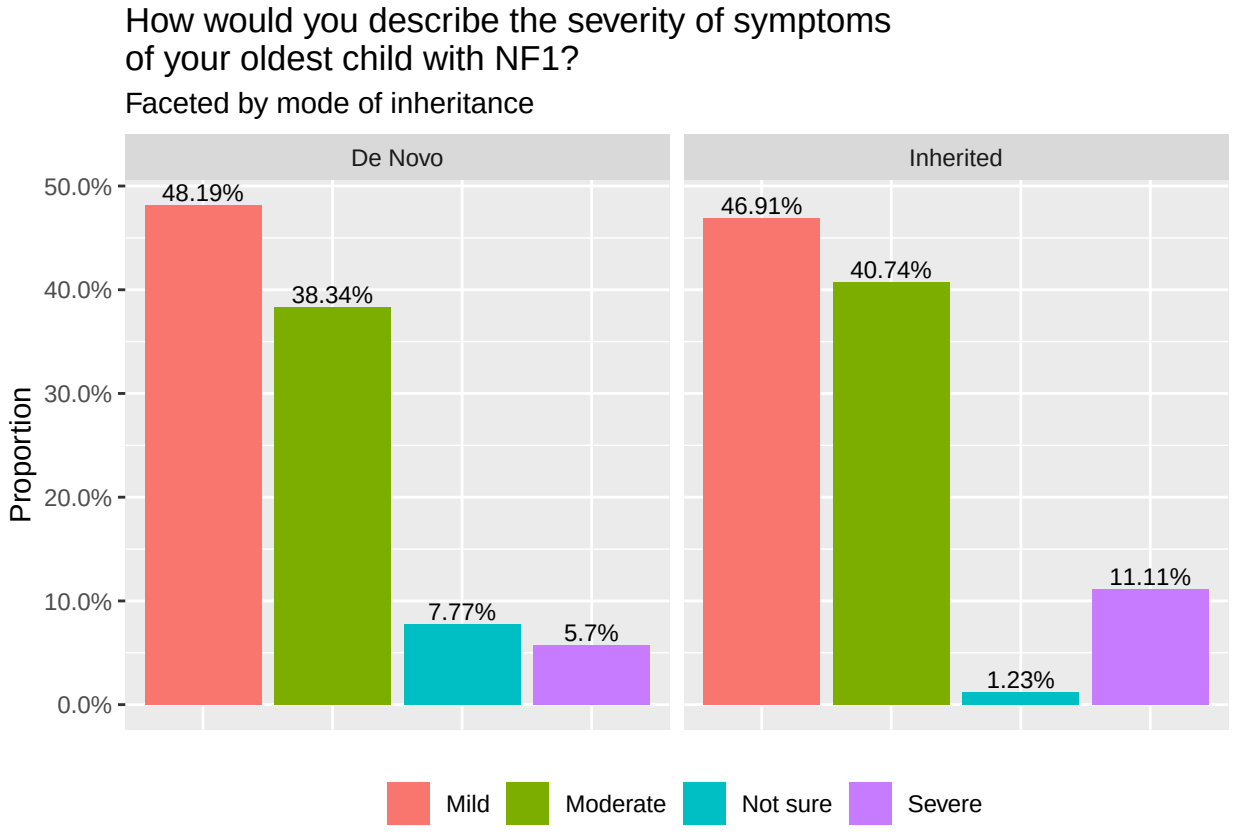
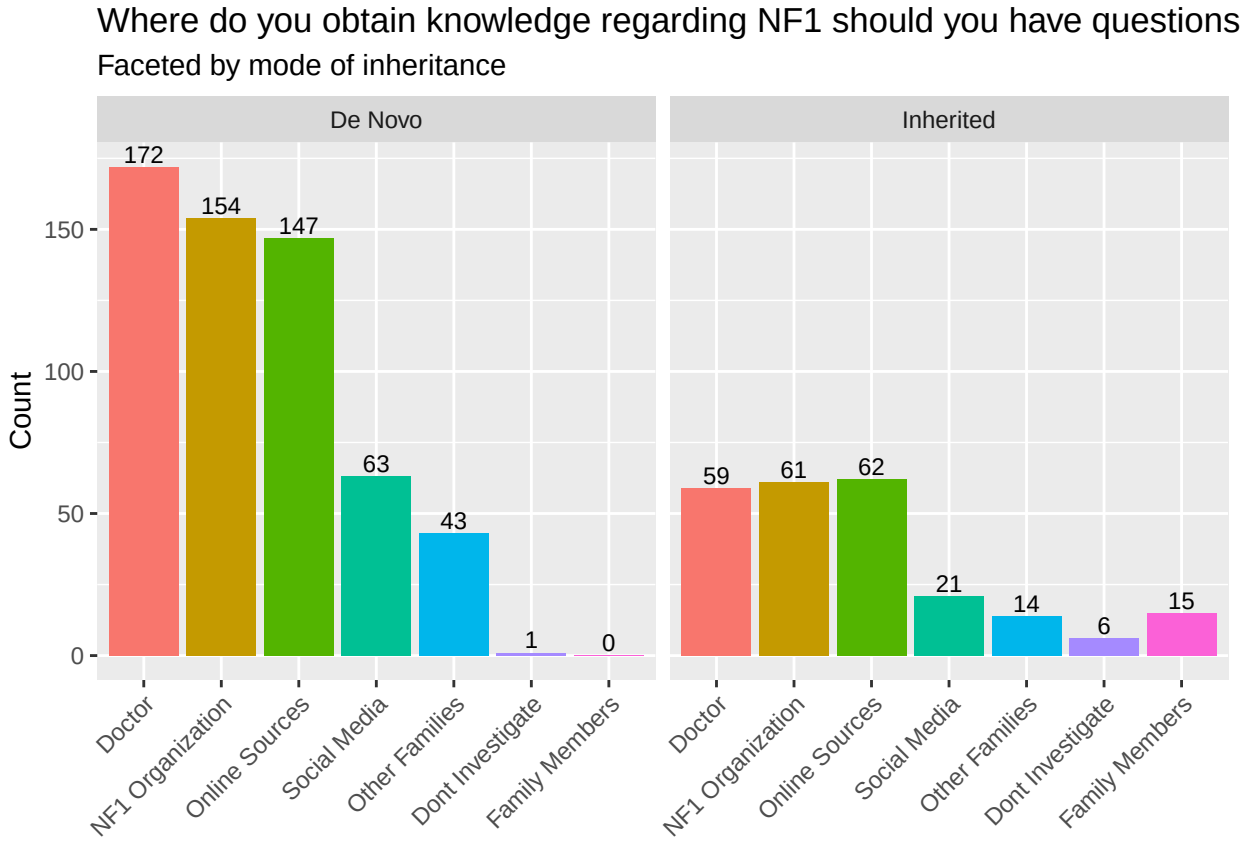

How would you describe the severity of symptoms of your oldest child with NF1?
Faceted by mode of inheritance
De Novo
Inherited
Proportion
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
48.19%
38.34%
7.77%
5.7%
46.91%
40.74%
1.23%
11.11%
Mild
Moderate
Not sure
Severe

Table 11: How would you describe the severity of symptoms of your oldest child with NF1?

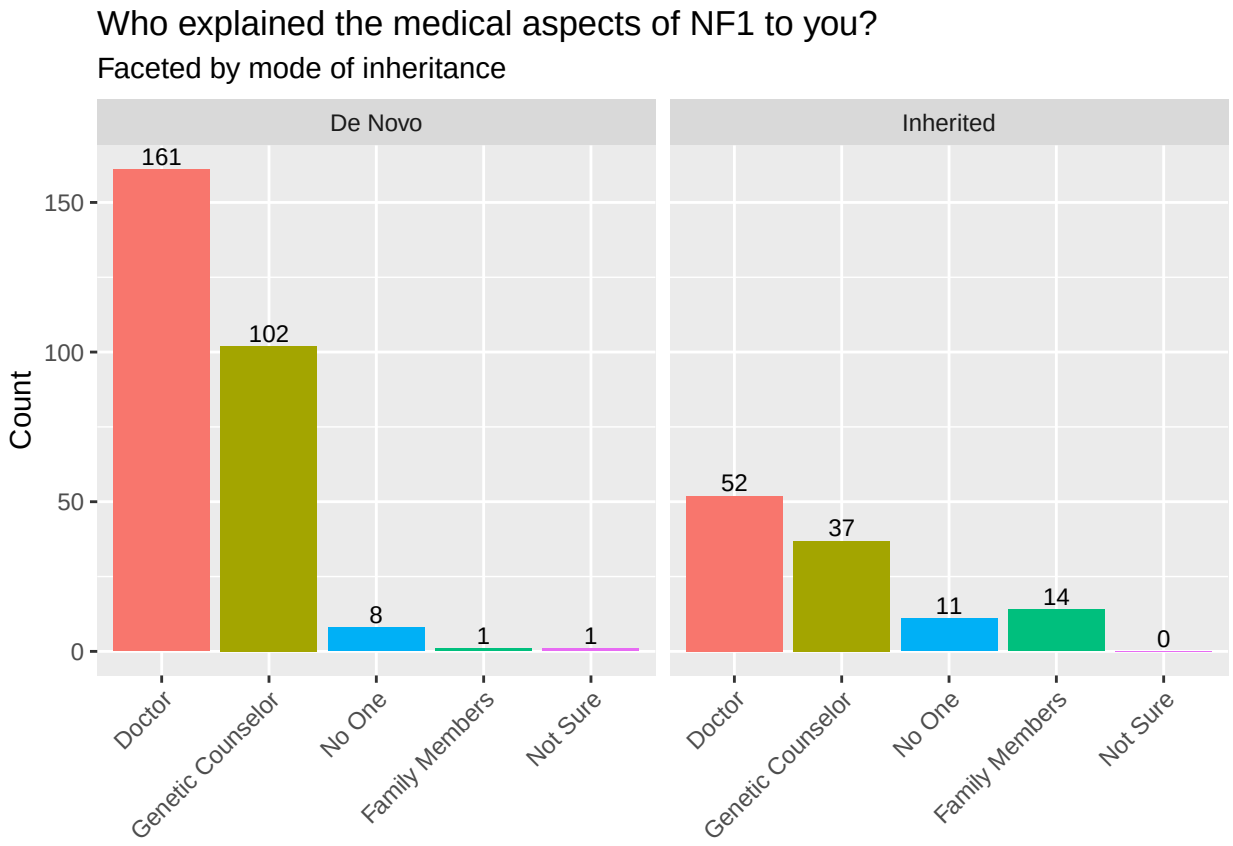
	De Novo	Inherited
Mild	93	38
Moderate	74	33
Not sure	15	1
Severe	11	9

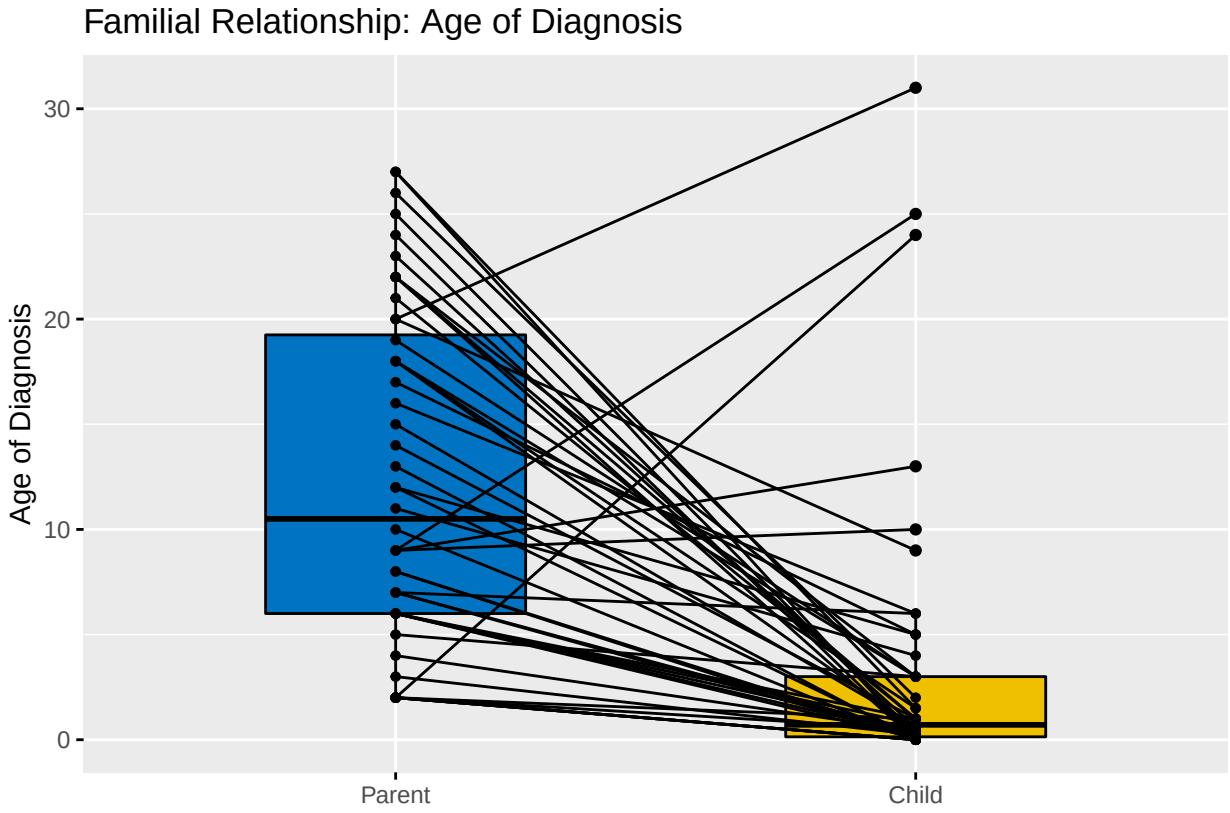
Similar to the "What specialists does your affected child see?" plot, it is important to note that the answers to the below question, "Where do you obtain knowledge regarding NF1 should you have questions?" are not mutually exclusive - parents who ask their doctor might also reference online sources.

```
## Warning in melt(display.df, "mode_of_inheritance"): The melt generic in
## data.table has been passed a tbl_df and will attempt to redirect to the relevant
## reshape2 method; please note that reshape2 is deprecated, and this redirection
## is now deprecated as well. To continue using melt methods from reshape2 while
## both libraries are attached, e.g. melt.list, you can prepend the namespace like
## reshape2::melt(display.df). In the next version, this warning will become an
## error.
```




```
## Warning in melt(display.df, "mode_of_inheritance"): The melt generic in
## data.table has been passed a tbl_df and will attempt to redirect to the relevant
## reshape2 method; please note that reshape2 is deprecated, and this redirection
## is now deprecated as well. To continue using melt methods from reshape2 while
## both libraries are attached, e.g. melt.list, you can prepend the namespace like
## reshape2::melt(display.df). In the next version, this warning will become an
## error.
```


Who explained the medical aspects of NF1 to you?
Faceted by mode of inheritance
De Novo
Inherited
Count
150
100
50
0
161
102
8
1
1
52
37
11
14
0
Doctor
Genetic Counselor
No One
Family Members
Not Sure


Familial Relationship: Age of Diagnosis
Age of Diagnosis
30
20
10
0
Parent
Child

References

[1] McDonald, J.H. (2014). Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland.

[2] Privitera, G. (2012). Statistics for the behavioral sciences. Thousand Oaks, CA: SAGE, p.367.