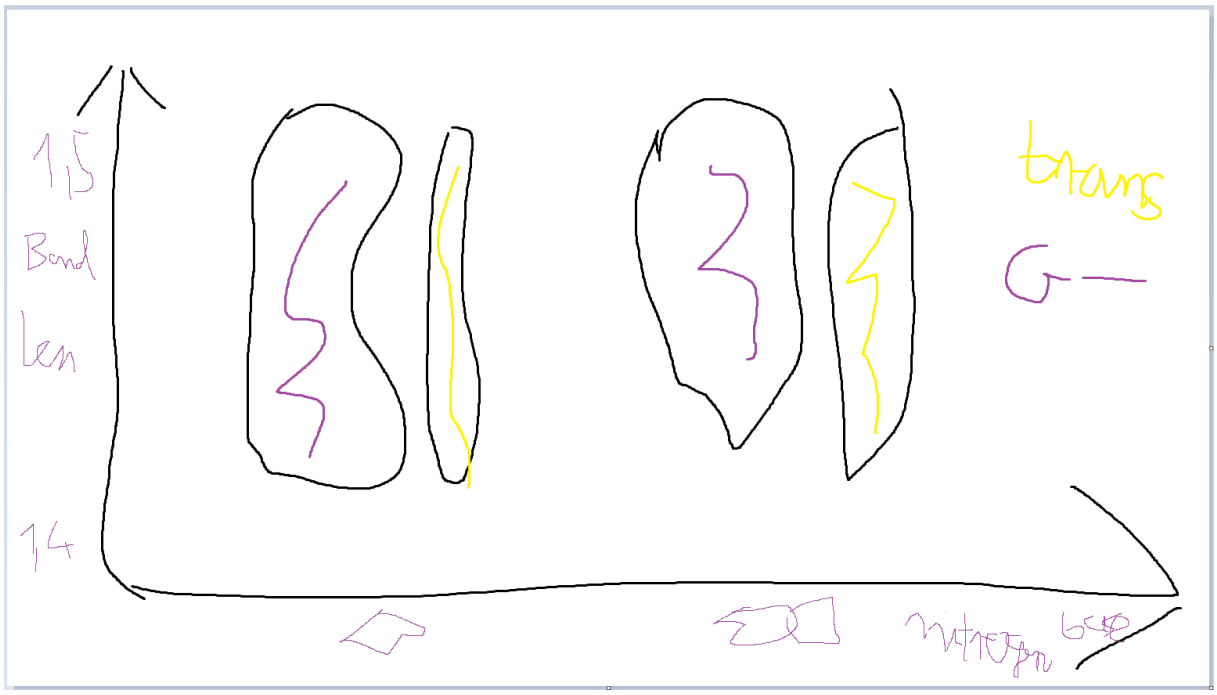
Data Visualization: Assignment 2

Student 1: Kacper Trębacz 145453

Student 2: Jan Gruszczyński 145464

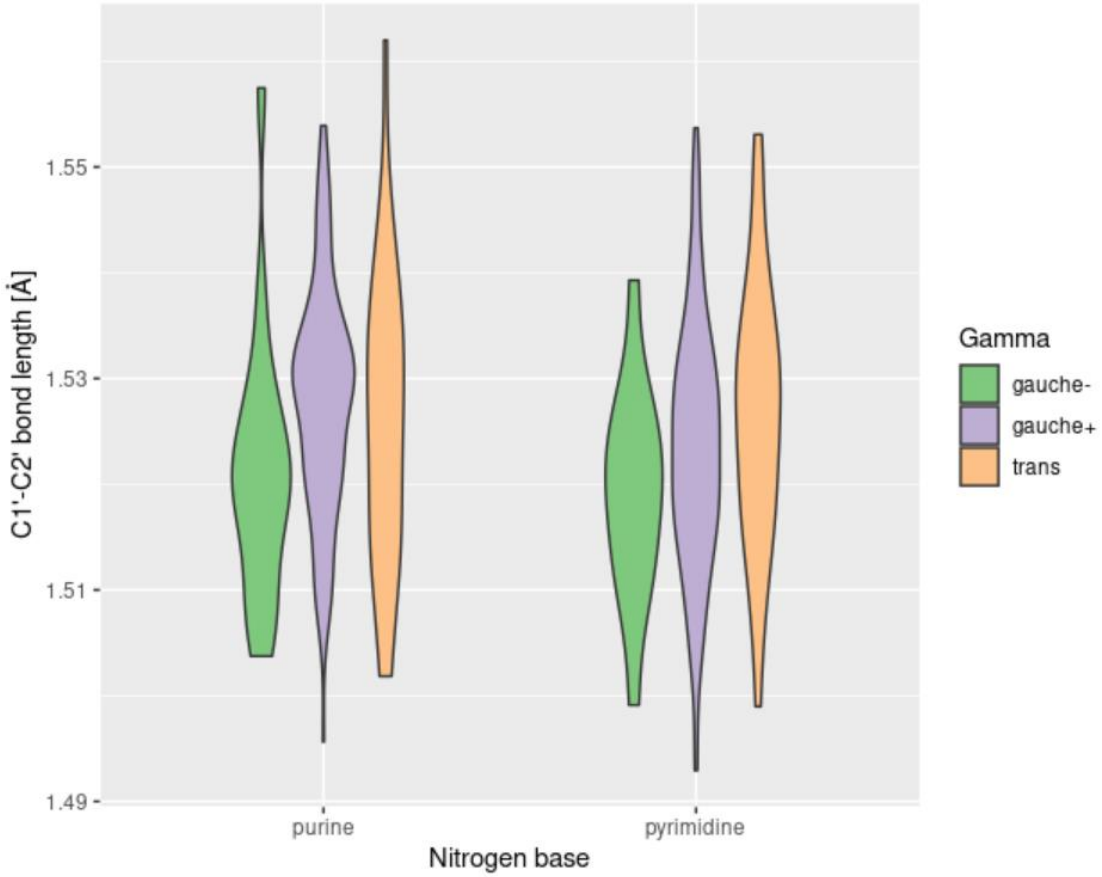
Three preliminary visualizations

Sketch 1



Implementation 1

Classic violin plot – nothing special to observe, beside the fact that the type of the nitrogen base has some effect on the bond's length.

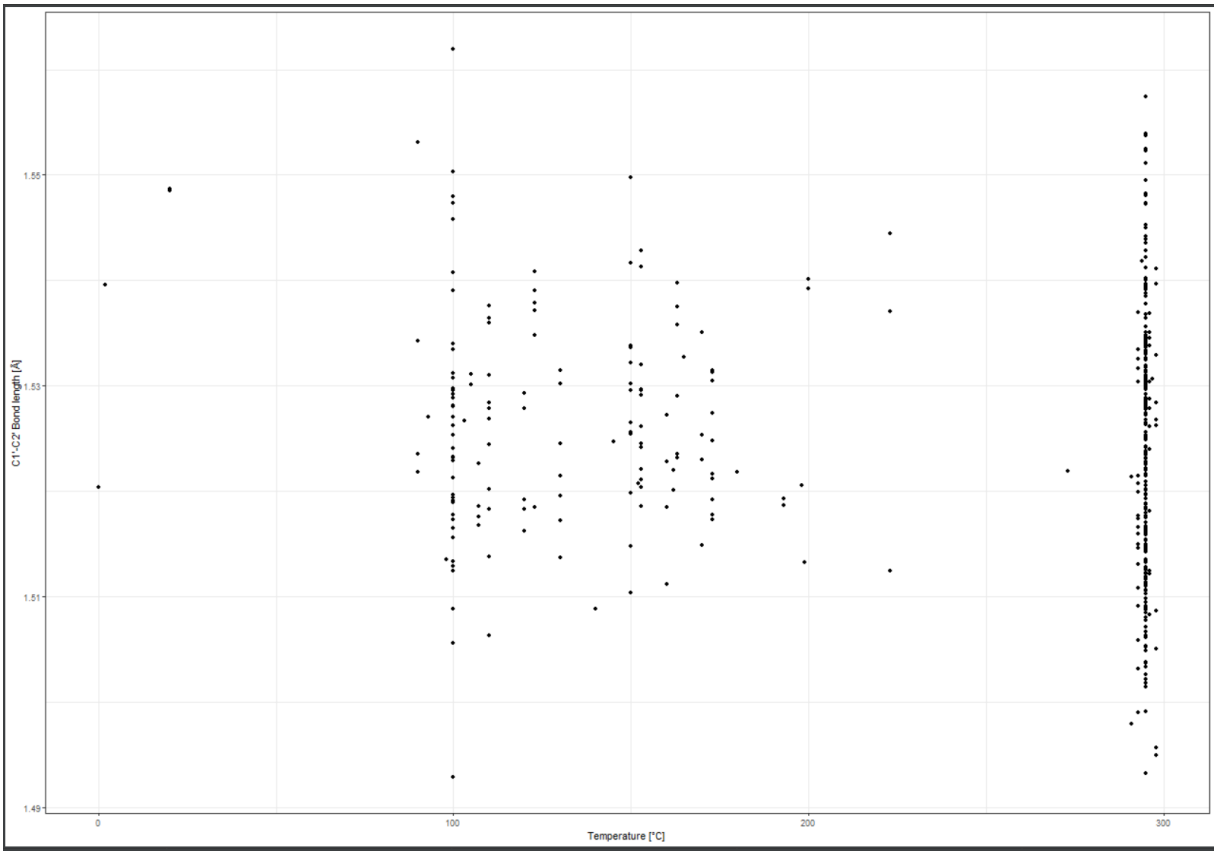


Implementation code 1

```
dodge <- position_dodge(width = 0.5)
ggplot(data, aes_string(x = "Base", y = "C1..C2.", fill = "Gamma")) +
  geom_violin(width = 0.5, position = dodge) +
  xlab("Nitrogen base") +
  ylab("C1'-C2' bond length [Å]") +
  scale_fill_brewer(type="qual", palette = 1)
```

Sketch 2 (optional)

Implementation 2



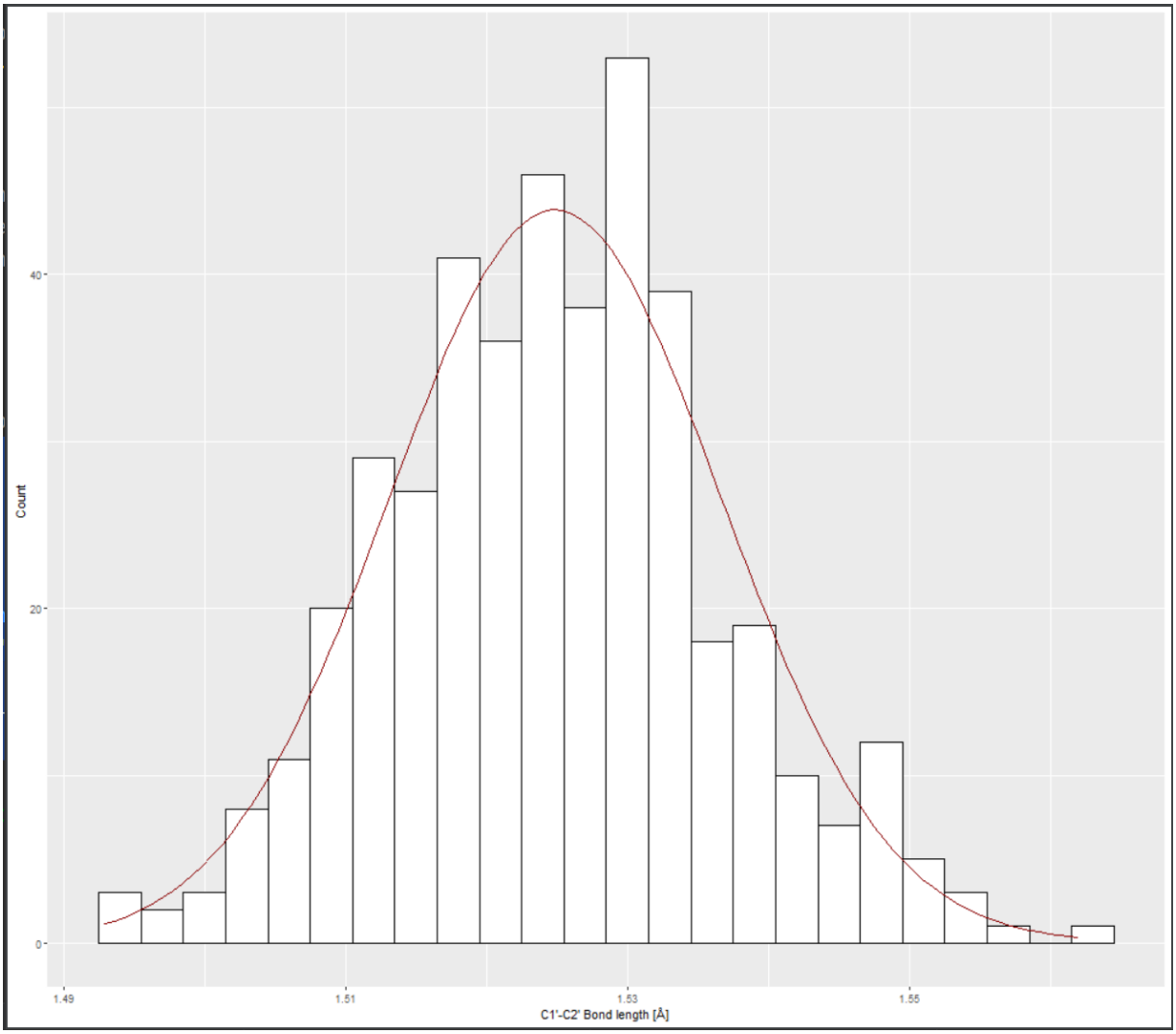
Shockingly for us, the temperature of measurement seems not to have an effect on given bond length. (We have checked every possible bond/angle)

Implementation code 2

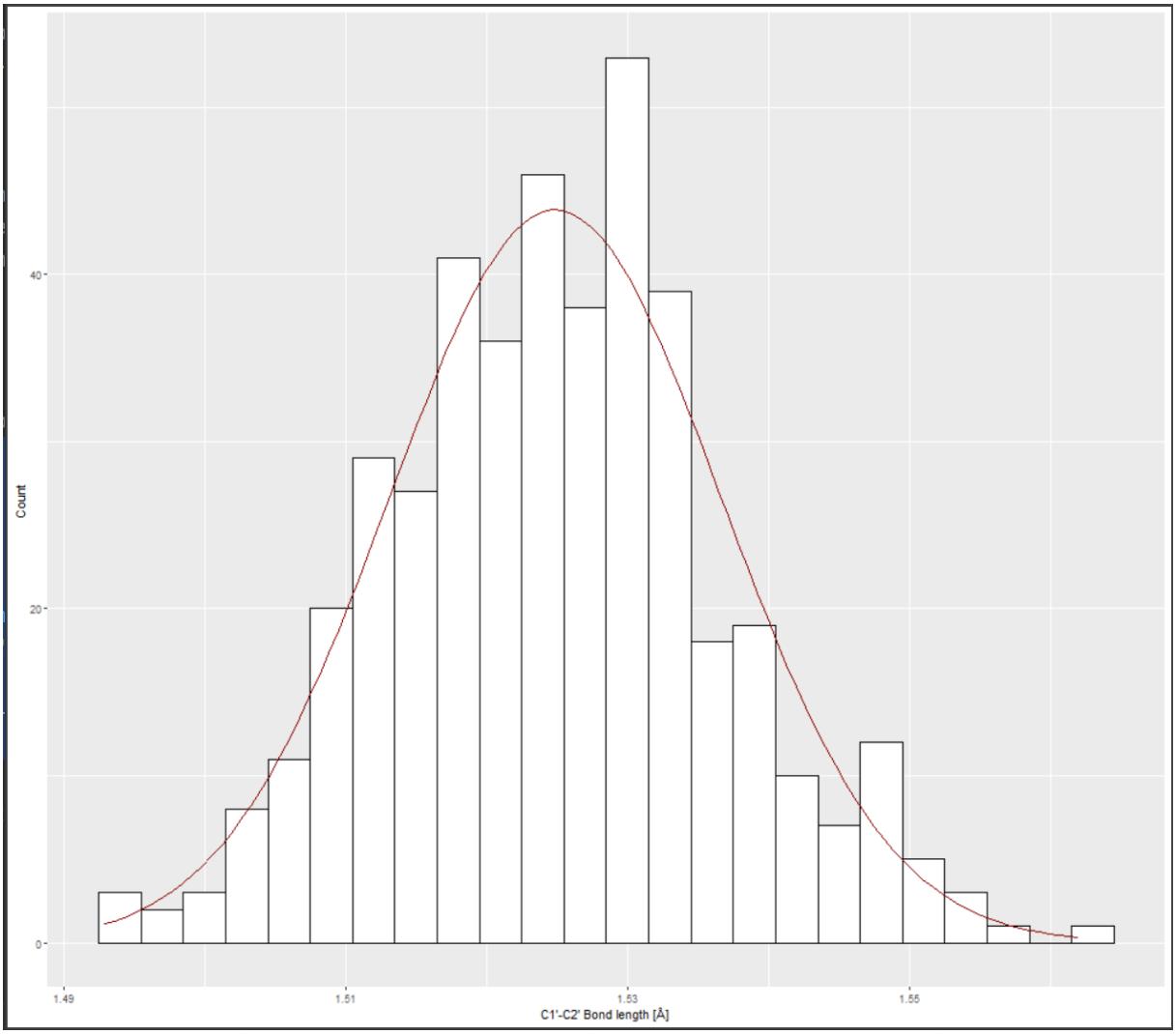
```r
#Prelimenary plot 2: temperature ~ bond length
ggplot(data = data, aes_string(x = "Temperature", y = "C1..C2.", group = 1)) +
  geom_point() +
  xlab( label: "Temperature [°C]") +
  ylab( label: "C1'-C2' Bond length [Å]") +
  theme_bw()

#for (name in c("C1..C2.", "C2..C3.", "C3..O3.", "C1..O4.",
#               "C2..O2.", "C4..C5.", "C4..O4.", "C5..O5.",
#               "C1..C2..C3.", "C1..C2..O2.", "C1..O4..C4.",
#               "C2..C1..O4.", "C2..C3..C4.", "C2..C3..O3.",
#               "C3..C2..O2.", "C3..C2..O2.", "C3..C4.", "C3..C4..C5.",
#               "C3..C4..O4.", "C4..C3..O3.", "C4..C5..O5.", "C5..C4..O4.")) {
#  print(ggplot(data = data, aes_string(x = "Temperature", y = name, group = 1)) +
#          geom_point() +
#          xlab("Temperature [°C]") +
#          ylab("C1-C2-C3 Bond angle [°]")) }
#Graphs shows that there seem to be no relation between temperature and bond angle/length
```

Sketch 3 (optional)

Implementation 3



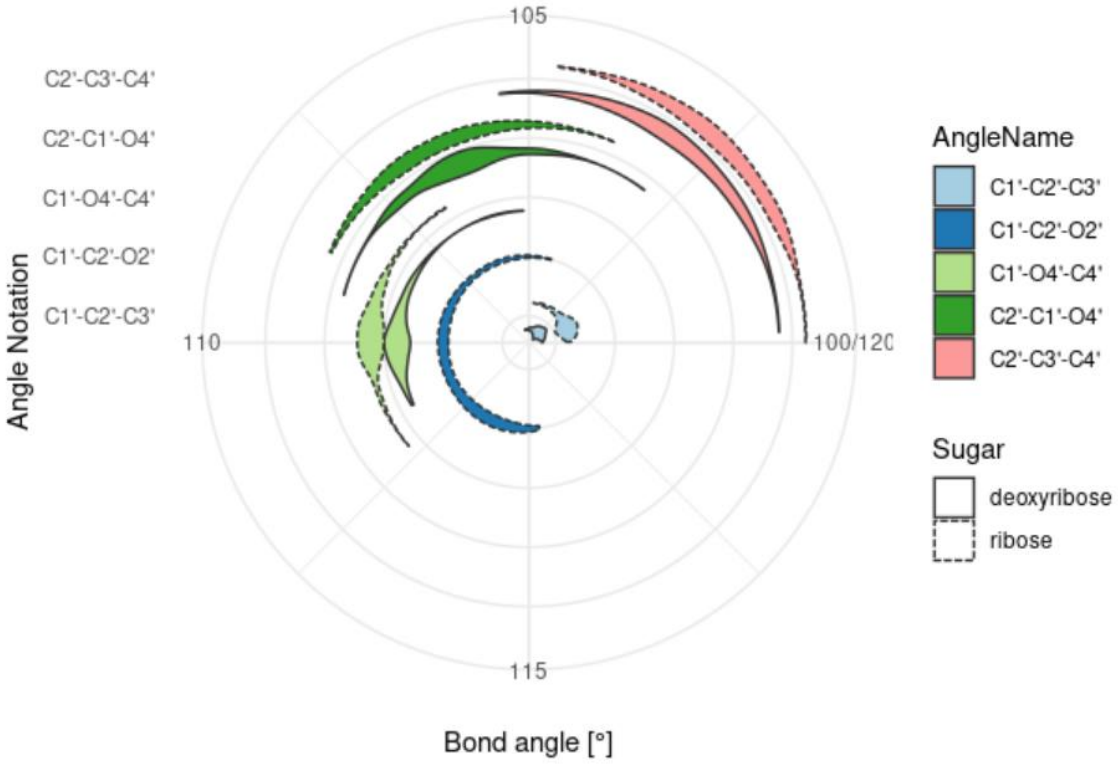
Classic histogram, red line shows that the measurements seem to be distributed according to normal distribution.



Implementation code 3

```
#Prelimenary plot 3 histogram + standard distribution
selected_data <- select(data, "C1..C2.")
values <- as.numeric(unlist(selected_data))
n <- length(values)
print(values)
ggplot(data, aes_string(x = "C1..C2."), binwidth = binwidth, n = n) +
  geom_histogram(color = "black", fill = "white", bins = 25, binwidth = binwidth) +
  stat_function(fun = function(x) dnorm(x, mean = mean(values), sd = sd(values)) * n * binwidth,
                color = "darkred", size = 0.5) +
  ylab( label: "Count") +
  xlab( label: "C1'-C2' Bond length [Å]")
```

Implementation 4



Experimenting with abstract visualization.

Implementation 4 code:

```r
```{r}
angles5 = read.csv("Angle5.csv", sep=",")
ggplot(angles5, aes(AngleName, Value, fill = AngleName, linetype = Sugar)) + geom_violin() +ylim(100,120) + coord_polar(theta = "y", direction = -1, start = -pi/2) +
theme_minimal() + scale_fill_brewer(type="qual", palette=3) +ylab("Bond angle [°]") + xlab("Angle Notation")
```
```

```
In [2]:
from plotnine import *
import pandas as pd
import numpy as np
```

```
In [3]:
biology = pd.read_csv("Sugars.csv")
```

```
In [4]:
array = np.array(biology.columns)
angles = list(biology.columns)
angles = [x for x in angles if x.count("-") == 2 and x.count("/")==0]
```

```
In [14]:
df = pd.DataFrame(columns=["AngleName","Sugar", "Value"])
for angle in angles:
    to_append = biology[["Sugar",angle]].set_index("Sugar").stack().reset_index().rename(columns={'level_1':'AngleName',
    df = pd.concat([df,to_append])
#df = df.set_index("AngleName")
df.describe()
df.to_csv("Angle.csv", index=False)
```

```
In [15]:
df = pd.DataFrame(columns=["AngleName","Sugar", "Value"])
for angle in angles[:5]:
    to_append = biology[["Sugar",angle]].set_index("Sugar").stack().reset_index().rename(columns={'level_1':'AngleName',
    df = pd.concat([df,to_append])
#df = df.set_index("AngleName")
df.describe()
df.to_csv("Angle5.csv", index=False)
```

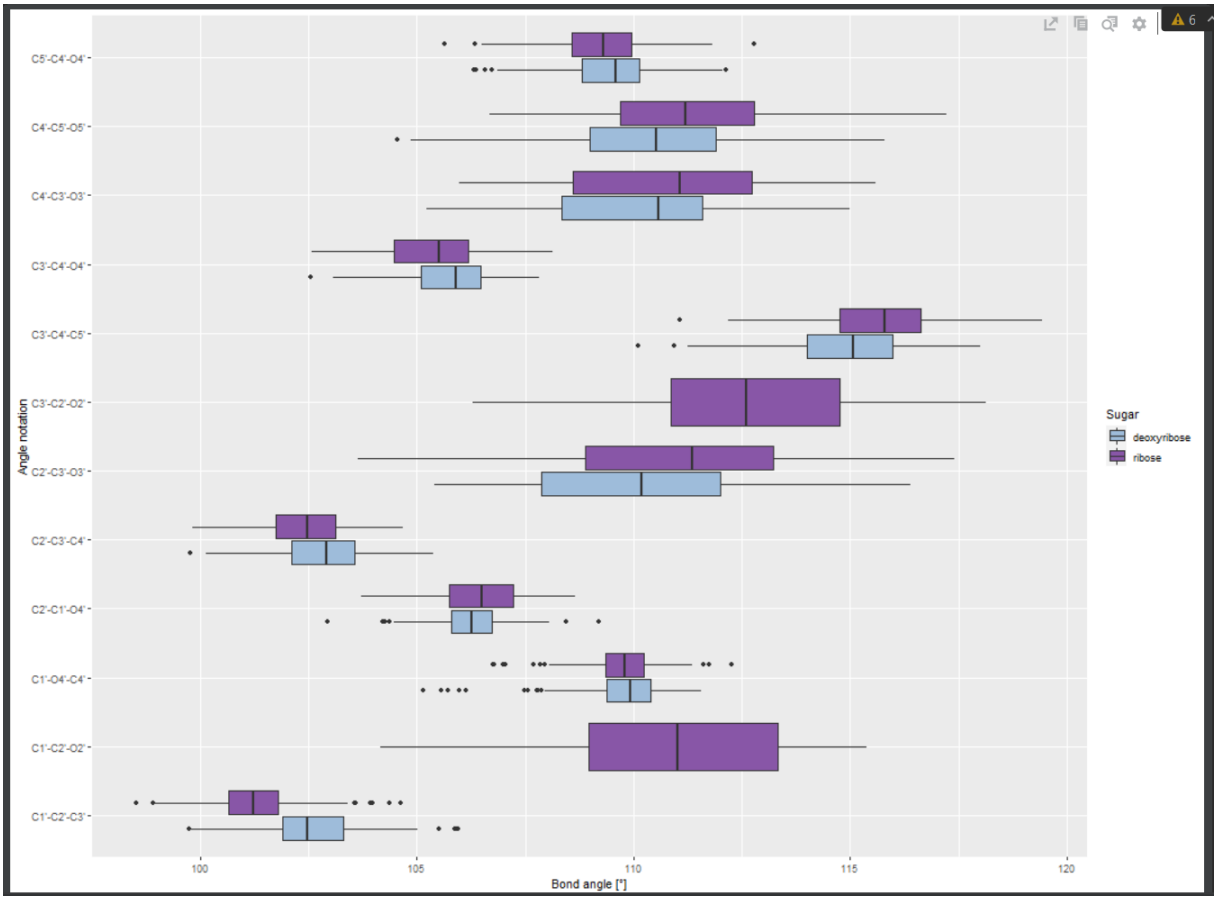
Implementation 5:

Implementation 5 code:

```r
```{r}
library(dplyr)
library(ggplot2)
library(GGally)
biology = read.csv("Sugars.csv", sep=",")

for( name in c("C1..C2.", "C2..C3.","C3..O3.", "C1..O4.", "C2..O2.","C4..C5.", "C4..O4.","C5..O5.","C1..C2..C3.", "C1..C2..O2.", "C1..O4..C4.", "C2..C1..O4.",
"C2..C3..C4.", "C2..C3..O3.", "C3..C2..O2.", "C3..C2..O2.", "C3..C4.", "C3..C4..C5.", "C3..C4..O4.", "C4..C3..O3.", "C4..C5..O5.", "C5..C4..O4.")){
  print(ggplot(biology, aes_string("Base",name, color=name)) + geom_violin(width = 0.5) )
}
```
```

Implementation 6:

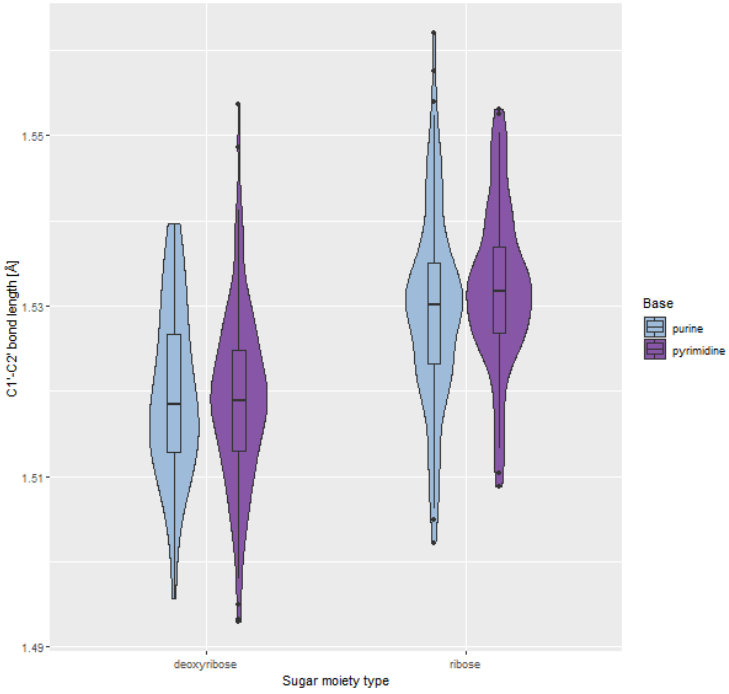


Implementation 6 code:

```r
angles = read.csv( file: "Angle.csv", sep=",")
#ggplot(angles, aes(Value, AngleName, fill = Sugar)) + geom_violin(width=1)
ggplot(angles, aes(Value, AngleName, fill = Sugar)) + geom_boxplot() + xlab( label: "Bond angle
 [°]") +ylab( label: "Angle notation") + scale_fill_manual(values= c("#9ebcda", "#8856a7"))
#ggplot(angles, aes(Value, AngleName, linetype = Sugar)) + geom_line() + coord_polar()
```

Selected final visualization

Implementation



This graph was chosen, as it seems to be the easiest to read and is full of information. The role of the atom of oxygen is shown, the effect of nitrogen base, box plots shows distribution and mean of the data.

Implementation code

```
dodge <- position_dodge(width = 0.5)
ggplot(data, aes_string(x = "Sugar", y = "C1..C2.", fill = "Base")) +
  geom_violin(width = 0.5, position = dodge) +
  geom_boxplot(width = 0.1, position = dodge) +
  theme_gray() +
  xlab( label: "Sugar moiety type" ) +
  ylab( label: "C1'-C2' bond length [Å]") +
  scale_fill_manual(values= c("#9ebcda", "#8856a7"))
```