

# 01\_Characterstics

2025-03-10

## Contents

Load . . . . .	1
Basic Exploration . . . . .	6

```
library(dplyr)
library(ggplot2)
library(reshape2)
```

## Load

Loading Pre-processed data provided

```
## load
load(here::here("Data", "ProcessedData", "processed_AnalysisData.Rdata"))
```

```
processed_data
```

```
## # A tibble: 14,575 x 484
##   fishNum dateSample dateTimeSample    dateProcessed species  spCode
##   <chr>    <date>      <dtm>          <date>        <chr>    <dbl>
## 1 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 2 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 3 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 4 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 5 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 6 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 7 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 8 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 9 LT001   2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## 10 LT001  2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout 81
## # i 14,565 more rows
## # i 478 more variables: totalLength <dbl>, forkLength <dbl>, weight <dbl>,
## #   girth <dbl>, dorsoLatHeight <dbl>, clipTag <chr>, sex <chr>, mat <dbl>,
## #   airbladderTotalLength <dbl>, airBladderWidth <dbl>, airbladderWeight <dbl>,
## #   airBladderWeightCond <dbl>, agingStructure <chr>, tissueSample <chr>,
## #   Region_name <chr>, FishTrack <chr>, MaxTSdiff <dbl>, Ping_time <chr>,
## #   deltaRange <dbl>, deltaMinAng <dbl>, deltaMajAng <dbl>, ...
```

```
## variable names in data
names(processed_data)
```

```
## [1] "fishNum"           "dateSample"
## [3] "dateTimeSample"    "dateProcessed"
## [5] "species"           "spCode"
## [7] "totalLength"       "forkLength"
## [9] "weight"            "girth"
```

## [11]	"dorsoLatHeight"	"clipTag"
## [13]	"sex"	"mat"
## [15]	"airbladderTotalLength"	"airBladderWidth"
## [17]	"airbladderWeight"	"airBladderWeightCond"
## [19]	"agingStructure"	"tissueSample"
## [21]	"Region_name"	"FishTrack"
## [23]	"MaxTSdiff"	"Ping_time"
## [25]	"deltaRange"	"deltaMinAng"
## [27]	"deltaMajAng"	"aspectAngle"
## [29]	"Target_range"	"Angle_minor_axis"
## [31]	"Angle_major_axis"	"Distance_minor_axis"
## [33]	"Distance_major_axis"	"StandDev_Angles_Minor_Axis"
## [35]	"StandDev_Angles_Major_Axis"	"Target_true_depth"
## [37]	"pingNumber"	"Ping_S"
## [39]	"Ping_E"	"Num_targets"
## [41]	"TS_mean"	"Target_range_mean"
## [43]	"Speed_4D_mean_unsmoothed"	"Fish_track_change_in_range"
## [45]	"Time_in_beam"	"Distance_3D_unsmoothed"
## [47]	"Thickness_mean"	"Exclude_below_line_range_mean"
## [49]	"Target_depth_mean"	"Target_depth_max"
## [51]	"Target_depth_min"	"Fish_track_change_in_depth"
## [53]	"Region_bottom_altitude_min"	"Region_bottom_altitude_max"
## [55]	"Region_bottom_altitude_mean"	"Region_top_altitude_min"
## [57]	"Region_top_altitude_max"	"Region_top_altitude_mean"
## [59]	"F45"	"F45.5"
## [61]	"F46"	"F46.5"
## [63]	"F47"	"F47.5"
## [65]	"F48"	"F48.5"
## [67]	"F49"	"F49.5"
## [69]	"F50"	"F50.5"
## [71]	"F51"	"F51.5"
## [73]	"F52"	"F52.5"
## [75]	"F53"	"F53.5"
## [77]	"F54"	"F54.5"
## [79]	"F55"	"F55.5"
## [81]	"F56"	"F56.5"
## [83]	"F57"	"F57.5"
## [85]	"F58"	"F58.5"
## [87]	"F59"	"F59.5"
## [89]	"F60"	"F60.5"
## [91]	"F61"	"F61.5"
## [93]	"F62"	"F62.5"
## [95]	"F63"	"F63.5"
## [97]	"F64"	"F64.5"
## [99]	"F65"	"F65.5"
## [101]	"F66"	"F66.5"
## [103]	"F67"	"F67.5"
## [105]	"F68"	"F68.5"
## [107]	"F69"	"F69.5"
## [109]	"F70"	"F70.5"
## [111]	"F71"	"F71.5"
## [113]	"F72"	"F72.5"
## [115]	"F73"	"F73.5"
## [117]	"F74"	"F74.5"

## [119]	"F75"	"F75.5"
## [121]	"F76"	"F76.5"
## [123]	"F77"	"F77.5"
## [125]	"F78"	"F78.5"
## [127]	"F79"	"F79.5"
## [129]	"F80"	"F80.5"
## [131]	"F81"	"F81.5"
## [133]	"F82"	"F82.5"
## [135]	"F83"	"F83.5"
## [137]	"F84"	"F84.5"
## [139]	"F85"	"F85.5"
## [141]	"F86"	"F86.5"
## [143]	"F87"	"F87.5"
## [145]	"F88"	"F88.5"
## [147]	"F89"	"F89.5"
## [149]	"F90"	"F90.5"
## [151]	"F91"	"F91.5"
## [153]	"F92"	"F92.5"
## [155]	"F93"	"F93.5"
## [157]	"F94"	"F94.5"
## [159]	"F95"	"F95.5"
## [161]	"F96"	"F96.5"
## [163]	"F97"	"F97.5"
## [165]	"F98"	"F98.5"
## [167]	"F99"	"F99.5"
## [169]	"F100"	"F100.5"
## [171]	"F101"	"F101.5"
## [173]	"F102"	"F102.5"
## [175]	"F103"	"F103.5"
## [177]	"F104"	"F104.5"
## [179]	"F105"	"F105.5"
## [181]	"F106"	"F106.5"
## [183]	"F107"	"F107.5"
## [185]	"F108"	"F108.5"
## [187]	"F109"	"F109.5"
## [189]	"F110"	"F110.5"
## [191]	"F111"	"F111.5"
## [193]	"F112"	"F112.5"
## [195]	"F113"	"F113.5"
## [197]	"F114"	"F114.5"
## [199]	"F115"	"F115.5"
## [201]	"F116"	"F116.5"
## [203]	"F117"	"F117.5"
## [205]	"F118"	"F118.5"
## [207]	"F119"	"F119.5"
## [209]	"F120"	"F120.5"
## [211]	"F121"	"F121.5"
## [213]	"F122"	"F122.5"
## [215]	"F123"	"F123.5"
## [217]	"F124"	"F124.5"
## [219]	"F125"	"F125.5"
## [221]	"F126"	"F126.5"
## [223]	"F127"	"F127.5"
## [225]	"F128"	"F128.5"

## [227]	"F129"	"F129.5"
## [229]	"F130"	"F130.5"
## [231]	"F131"	"F131.5"
## [233]	"F132"	"F132.5"
## [235]	"F133"	"F133.5"
## [237]	"F134"	"F134.5"
## [239]	"F135"	"F135.5"
## [241]	"F136"	"F136.5"
## [243]	"F137"	"F137.5"
## [245]	"F138"	"F138.5"
## [247]	"F139"	"F139.5"
## [249]	"F140"	"F140.5"
## [251]	"F141"	"F141.5"
## [253]	"F142"	"F142.5"
## [255]	"F143"	"F143.5"
## [257]	"F144"	"F144.5"
## [259]	"F145"	"F145.5"
## [261]	"F146"	"F146.5"
## [263]	"F147"	"F147.5"
## [265]	"F148"	"F148.5"
## [267]	"F149"	"F149.5"
## [269]	"F150"	"F150.5"
## [271]	"F151"	"F151.5"
## [273]	"F152"	"F152.5"
## [275]	"F153"	"F153.5"
## [277]	"F154"	"F154.5"
## [279]	"F155"	"F155.5"
## [281]	"F156"	"F156.5"
## [283]	"F157"	"F157.5"
## [285]	"F158"	"F158.5"
## [287]	"F159"	"F159.5"
## [289]	"F160"	"F160.5"
## [291]	"F161"	"F161.5"
## [293]	"F162"	"F162.5"
## [295]	"F163"	"F163.5"
## [297]	"F164"	"F164.5"
## [299]	"F165"	"F165.5"
## [301]	"F166"	"F166.5"
## [303]	"F167"	"F167.5"
## [305]	"F168"	"F168.5"
## [307]	"F169"	"F169.5"
## [309]	"F170"	"F173"
## [311]	"F173.5"	"F174"
## [313]	"F174.5"	"F175"
## [315]	"F175.5"	"F176"
## [317]	"F176.5"	"F177"
## [319]	"F177.5"	"F178"
## [321]	"F178.5"	"F179"
## [323]	"F179.5"	"F180"
## [325]	"F180.5"	"F181"
## [327]	"F181.5"	"F182"
## [329]	"F182.5"	"F183"
## [331]	"F183.5"	"F184"
## [333]	"F184.5"	"F185"

## [335]	"F185.5"	"F186"
## [337]	"F186.5"	"F187"
## [339]	"F187.5"	"F188"
## [341]	"F188.5"	"F189"
## [343]	"F189.5"	"F190"
## [345]	"F190.5"	"F191"
## [347]	"F191.5"	"F192"
## [349]	"F192.5"	"F193"
## [351]	"F193.5"	"F194"
## [353]	"F194.5"	"F195"
## [355]	"F195.5"	"F196"
## [357]	"F196.5"	"F197"
## [359]	"F197.5"	"F198"
## [361]	"F198.5"	"F199"
## [363]	"F199.5"	"F200"
## [365]	"F200.5"	"F201"
## [367]	"F201.5"	"F202"
## [369]	"F202.5"	"F203"
## [371]	"F203.5"	"F204"
## [373]	"F204.5"	"F205"
## [375]	"F205.5"	"F206"
## [377]	"F206.5"	"F207"
## [379]	"F207.5"	"F208"
## [381]	"F208.5"	"F209"
## [383]	"F209.5"	"F210"
## [385]	"F210.5"	"F211"
## [387]	"F211.5"	"F212"
## [389]	"F212.5"	"F213"
## [391]	"F213.5"	"F214"
## [393]	"F214.5"	"F215"
## [395]	"F215.5"	"F216"
## [397]	"F216.5"	"F217"
## [399]	"F217.5"	"F218"
## [401]	"F218.5"	"F219"
## [403]	"F219.5"	"F220"
## [405]	"F220.5"	"F221"
## [407]	"F221.5"	"F222"
## [409]	"F222.5"	"F223"
## [411]	"F223.5"	"F224"
## [413]	"F224.5"	"F225"
## [415]	"F225.5"	"F226"
## [417]	"F226.5"	"F227"
## [419]	"F227.5"	"F228"
## [421]	"F228.5"	"F229"
## [423]	"F229.5"	"F230"
## [425]	"F230.5"	"F231"
## [427]	"F231.5"	"F232"
## [429]	"F232.5"	"F233"
## [431]	"F233.5"	"F234"
## [433]	"F234.5"	"F235"
## [435]	"F235.5"	"F236"
## [437]	"F236.5"	"F237"
## [439]	"F237.5"	"F238"
## [441]	"F238.5"	"F239"

```
## [443] "F239.5"          "F240"
## [445] "F240.5"          "F241"
## [447] "F241.5"          "F242"
## [449] "F242.5"          "F243"
## [451] "F243.5"          "F244"
## [453] "F244.5"          "F245"
## [455] "F245.5"          "F246"
## [457] "F246.5"          "F247"
## [459] "F247.5"          "F248"
## [461] "F248.5"          "F249"
## [463] "F249.5"          "F250"
## [465] "F250.5"          "F251"
## [467] "F251.5"          "F252"
## [469] "F252.5"          "F253"
## [471] "F253.5"          "F254"
## [473] "F254.5"          "F255"
## [475] "F255.5"          "F256"
## [477] "F256.5"          "F257"
## [479] "F257.5"          "F258"
## [481] "F258.5"          "F259"
## [483] "F259.5"          "F260"
```

## Basic Exploration

### Basic Characteristics

Number of total fish species.

```
unique(processed_data$species)
```

```
## [1] "lakeTrout"      "lakeWhitefish"  "smallmouthBass"
```

Number of fish sample in each fish species.

```
## "1" been considered as Male
## "2" been considered as Female
(
  processed_data
  |> group_by(species)
  |> summarize(n = n_distinct(fishNum),
              M = n_distinct(fishNum[sex == "1"]),
              F = n_distinct(fishNum[sex == "2"]))
)
```

```
## # A tibble: 3 x 4
##   species      n      M      F
##   <chr>    <int> <int> <int>
## 1 lakeTrout    21      8    13
## 2 lakeWhitefish 11      4      6
## 3 smallmouthBass 12      8      4
```

Create a basic characteristics table

```
character_data <- (
  processed_data
  |> select(1:58)
  |> group_by(species, fishNum)
  |> slice(1)
```

```
|> ungroup()
)
character_data

## # A tibble: 44 x 58
##   fishNum dateSample dateTimeSample   dateProcessed species  spCode
##   <chr>    <date>      <dtm>          <date>        <chr>    <dbl>
## 1 LT001    2022-07-21 2022-07-21 16:56:00 2022-07-27 lakeTrout    81
## 2 LT002    2022-07-22 2022-07-22 11:08:00 2022-07-27 lakeTrout    81
## 3 LT003    2022-07-22 2022-07-22 12:06:00 2022-07-27 lakeTrout    81
## 4 LT004    2022-07-22 2022-07-22 17:31:00 2022-07-27 lakeTrout    81
## 5 LT005    2022-07-23 2022-07-23 10:28:00 2022-07-27 lakeTrout    81
## 6 LT006    2022-07-23 2022-07-23 18:19:00 2022-07-26 lakeTrout    81
## 7 LT007    2022-07-23 2022-07-23 19:11:00 2022-07-27 lakeTrout    81
## 8 LT008    2022-07-26 2022-07-26 09:32:00 2022-07-26 lakeTrout    81
## 9 LT009    2022-07-26 2022-07-26 10:56:00 2022-07-26 lakeTrout    81
## 10 LT010   2022-07-26 2022-07-26 11:37:00 2022-07-26 lakeTrout    81
## # i 34 more rows
## # i 52 more variables: totalLength <dbl>, forkLength <dbl>, weight <dbl>,
## #   girth <dbl>, dorsoLatHeight <dbl>, clipTag <chr>, sex <chr>, mat <dbl>,
## #   airbladderTotalLength <dbl>, airBladderWidth <dbl>, airbladderWeight <dbl>,
## #   airBladderWeightCond <dbl>, agingStructure <chr>, tissueSample <chr>,
## #   Region_name <chr>, FishTrack <chr>, MaxTSdiff <dbl>, Ping_time <chr>,
## #   deltaRange <dbl>, deltaMinAng <dbl>, deltaMajAng <dbl>, ...
```

Separate into three dataset for each species

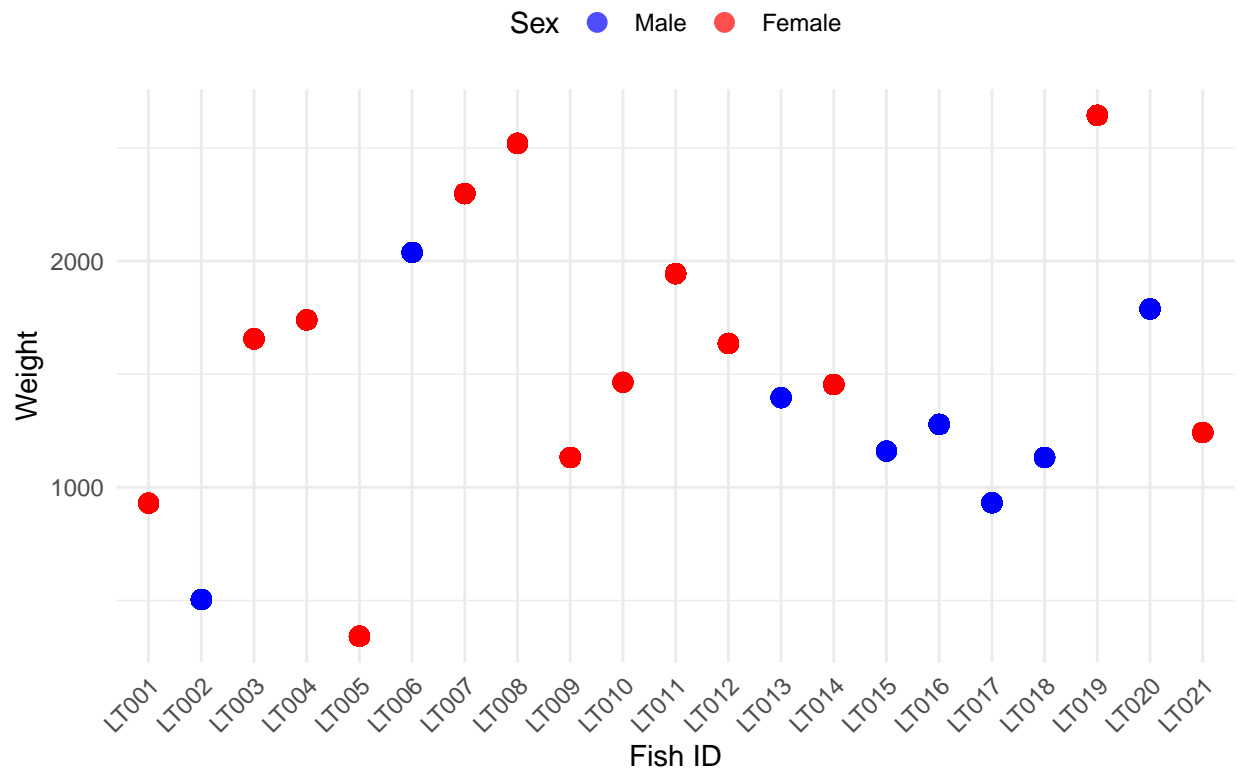
```
LakeTrout <- processed_data[processed_data$species == "lakeTrout", ]
LakeWhiteFish <- processed_data[processed_data$species == "lakeWhitefish", ]
SmallmouthBass <- processed_data[processed_data$species == "smallmouthBass", ]
```

## Weight

Individual fish weight in each species.

```
## LakeTrout, LakeWhiteFish, SmallmouthBass
weight_data = LakeTrout
(
  ggplot(weight_data, aes(x = fishNum, y = weight))
  + geom_point(aes(color = sex), size = 3, alpha = 0.7)
  + geom_line(aes(group = fishNum), alpha = 0.3)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
                        labels = c("1" = "Male", "2" = "Female"),
                        name = "Sex")
  + labs(
    title = paste0("Indiviaul Weight distribution in ", weight_data$species),
    x = "Fish ID",
    y = "Weight"
  )
  + theme_minimal()
  + theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "top"
  )
)
```

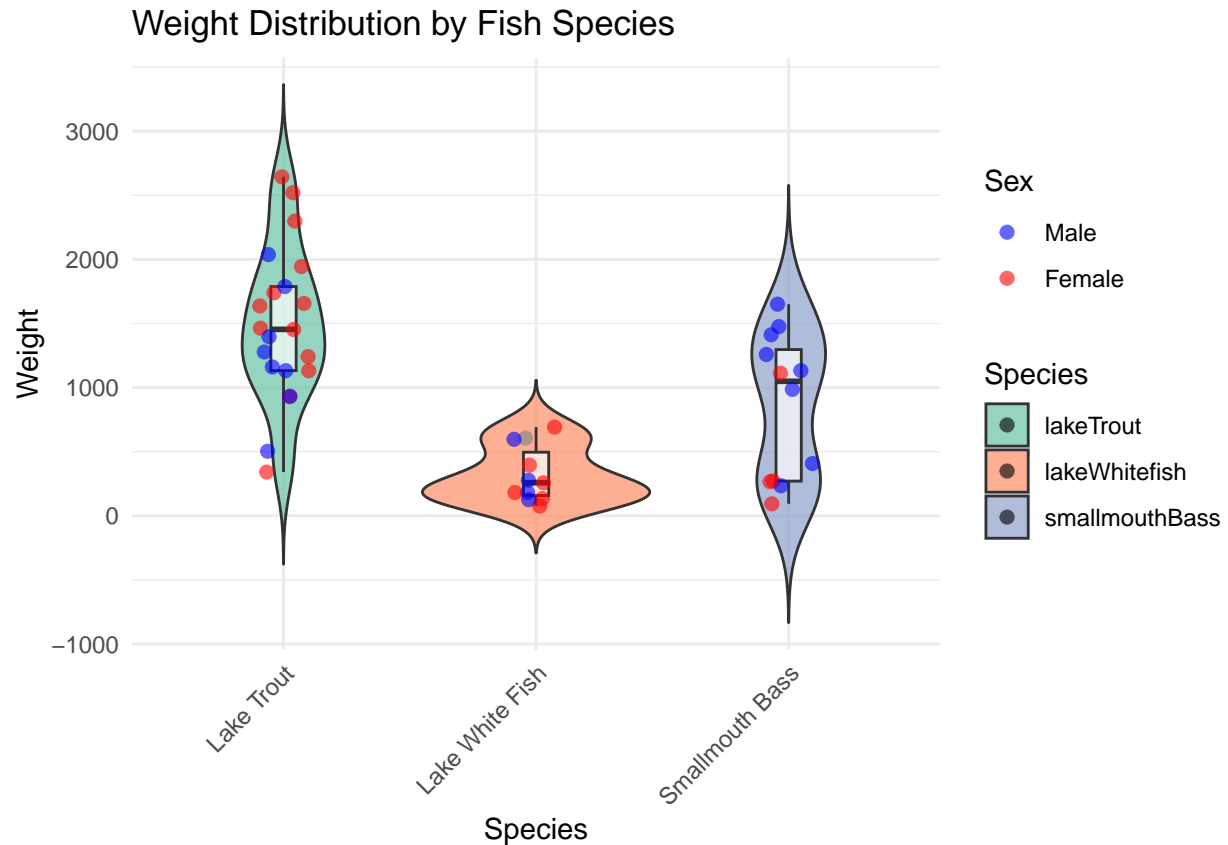
## Indiviaul Weight distribution in lakeTrout



Weight distribution comparion between three species.

```
(
  ggplot(character_data, aes(x = species, y = weight, fill = species))
  + geom_violin(alpha = 0.7, trim = FALSE)
  + geom_boxplot(width = 0.1, fill = "white", alpha = 0.7)
  + geom_jitter(aes(color = factor(sex)), width = 0.1, alpha = 0.6, size = 2)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
    labels = c("1" = "Male", "2" = "Female"),
    name = "Sex")
  + scale_fill_brewer(palette = "Set2")
  + scale_x_discrete(labels = c("lakeTrout" = "Lake Trout",
    "lakeWhitefish" = "Lake White Fish",
    "smallmouthBass" = "Smallmouth Bass"))
  + labs(
    title = "Weight Distribution by Fish Species",
    x = "Species",
    y = "Weight",
    fill = "Species"
  )
  + theme_minimal()
  + theme(
    legend.position = "right",
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
)
```



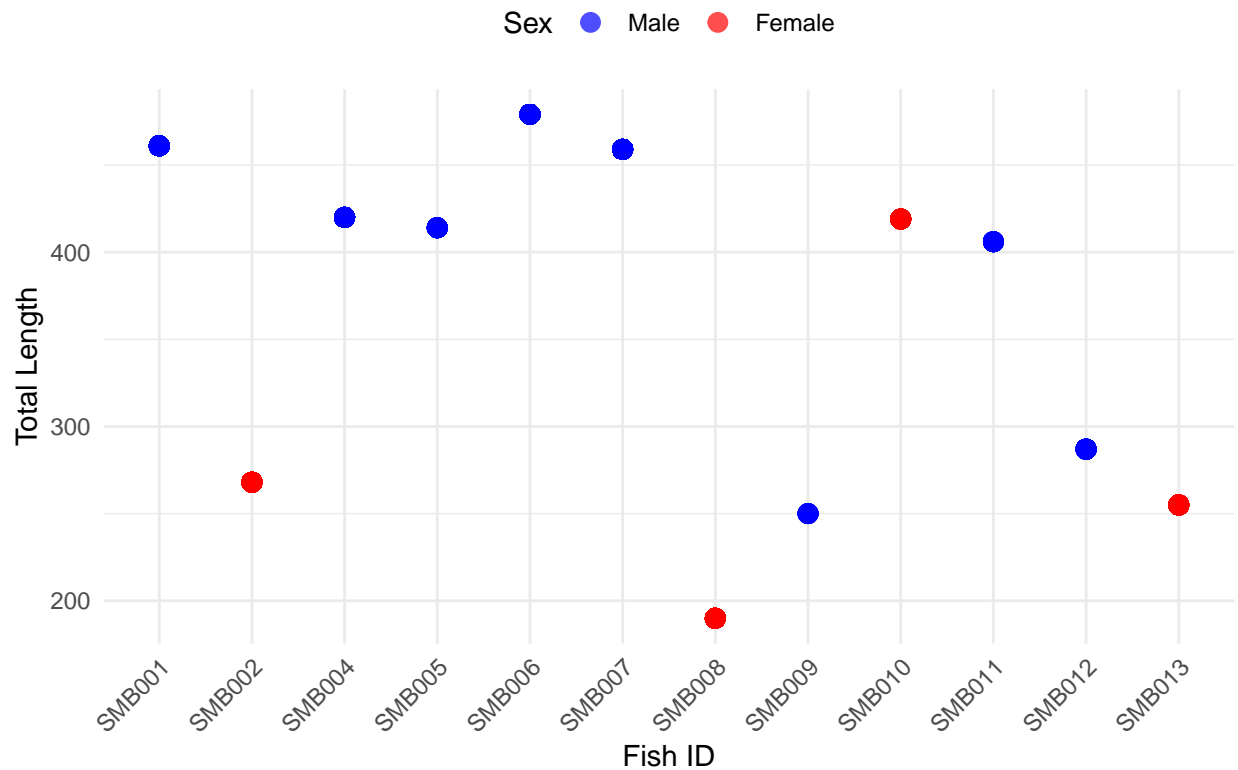


### Total Length

Individual fish total length in each species.

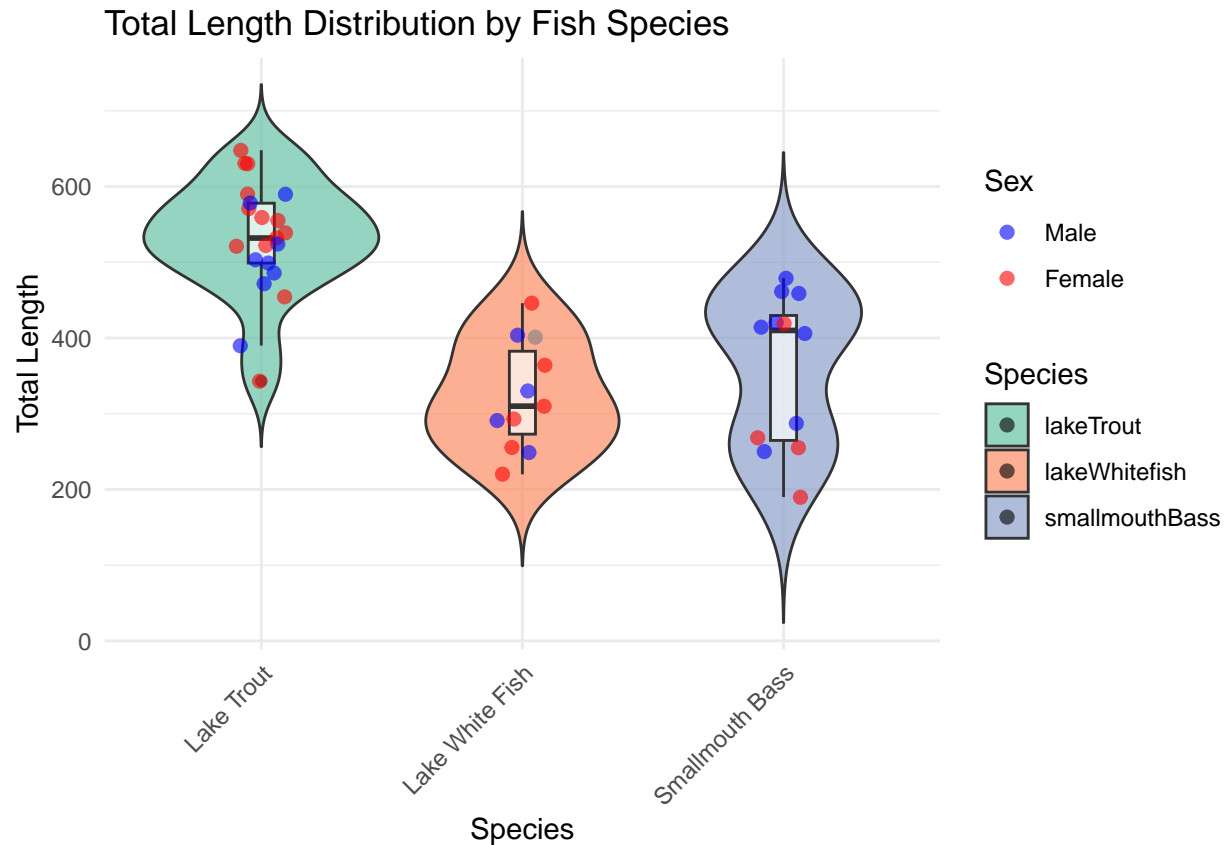
```
## LakeTrout, LakeWhiteFish, SmallmouthBass
length_data = SmallmouthBass
(
  ggplot(length_data, aes(x = fishNum, y = totalLength))
  + geom_point(aes(color = sex), size = 3, alpha = 0.7)
  + geom_line(aes(group = fishNum), alpha = 0.3)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
                        labels = c("1" = "Male", "2" = "Female"),
                        name = "Sex")
  + labs(
    title = paste0("Individual Total Length distribution in ", length_data$species),
    x = "Fish ID",
    y = "Total Length"
  )
  + theme_minimal()
  + theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "top"
  )
)
```

## Individual Total Length distribution in smallmouthBass



Total Length distribution comparison between three species.

```
(
  ggplot(character_data, aes(x = species, y = totalLength, fill = species))
  + geom_violin(alpha = 0.7, trim = FALSE)
  + geom_boxplot(width = 0.1, fill = "white", alpha = 0.7)
  + geom_jitter(aes(color = factor(sex)), width = 0.1, alpha = 0.6, size = 2)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
                        labels = c("1" = "Male", "2" = "Female"),
                        name = "Sex")
  + scale_fill_brewer(palette = "Set2")
  + scale_x_discrete(labels = c("lakeTrout" = "Lake Trout",
                                "lakeWhitefish" = "Lake White Fish",
                                "smallmouthBass" = "Smallmouth Bass"))
  + labs(
    title = "Total Length Distribution by Fish Species",
    x = "Species",
    y = "Total Length",
    fill = "Species"
  )
  + theme_minimal()
  + theme(
    legend.position = "right",
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
)
```



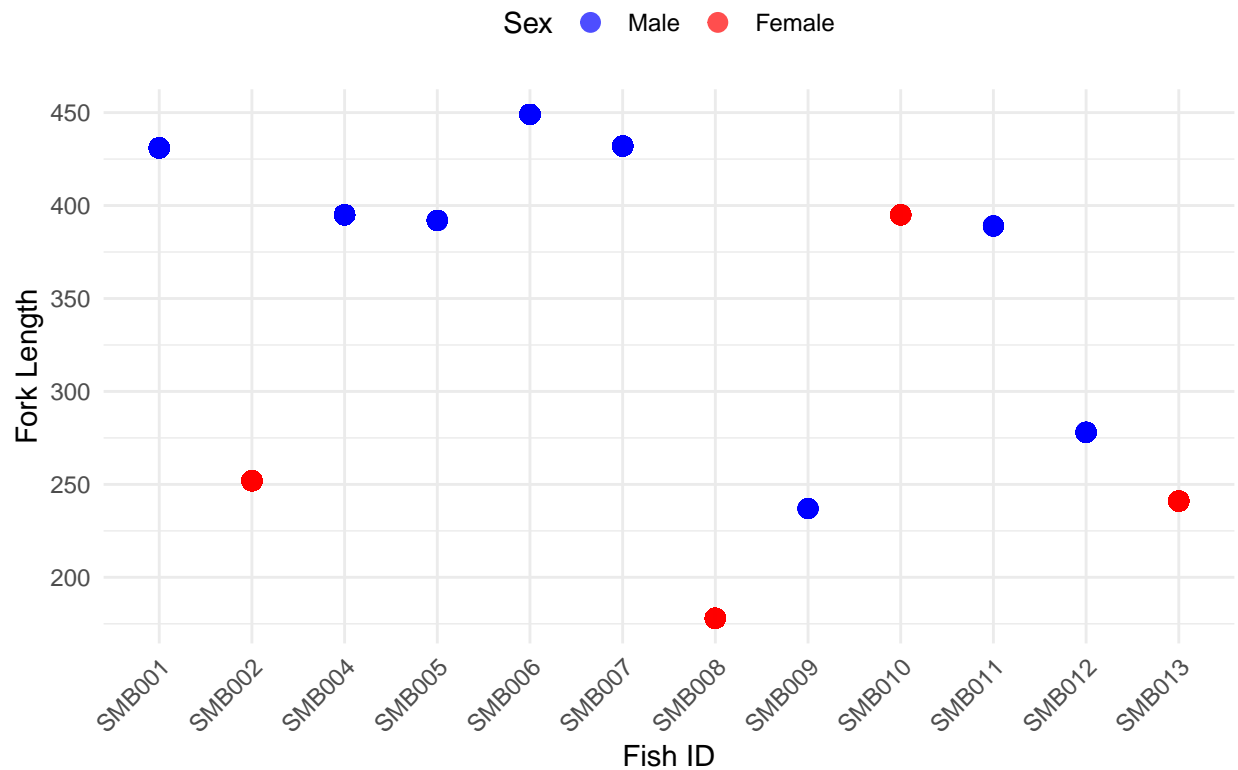
## Fork Length

Individual fish fork length in each species.

```
## LakeTrout, LakeWhiteFish, SmallmouthBass
length_data = SmallmouthBass
(
  ggplot(length_data, aes(x = fishNum, y = forkLength))
  + geom_point(aes(color = sex), size = 3, alpha = 0.7)
  + geom_line(aes(group = fishNum), alpha = 0.3)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
                        labels = c("1" = "Male", "2" = "Female"),
                        name = "Sex")

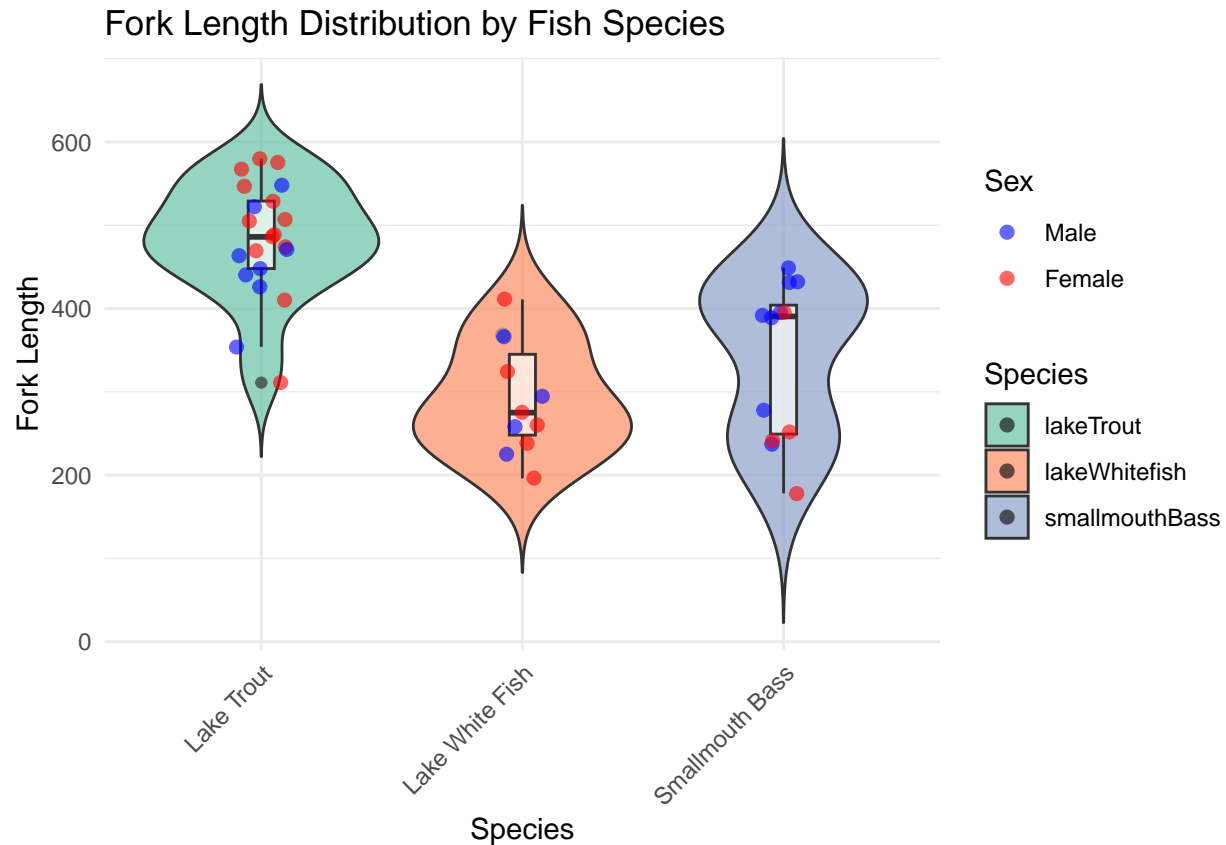
  + labs(
    title = paste0("Individual Fork Length distribution in ", length_data$species),
    x = "Fish ID",
    y = "Fork Length"
  )
  + theme_minimal()
  + theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "top"
  )
)
```

## Individual Fork Length distribution in smallmouthBass



Total Length distribution comparison between three species.

```
(
  ggplot(character_data, aes(x = species, y = forkLength, fill = species))
  + geom_violin(alpha = 0.7, trim = FALSE)
  + geom_boxplot(width = 0.1, fill = "white", alpha = 0.7)
  + geom_jitter(aes(color = factor(sex)), width = 0.1, alpha = 0.6, size = 2)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
                        labels = c("1" = "Male", "2" = "Female"),
                        name = "Sex")
  + scale_fill_brewer(palette = "Set2")
  + scale_x_discrete(labels = c("lakeTrout" = "Lake Trout",
                                "lakeWhitefish" = "Lake White Fish",
                                "smallmouthBass" = "Smallmouth Bass"))
  + labs(
    title = "Fork Length Distribution by Fish Species",
    x = "Species",
    y = "Fork Length",
    fill = "Species"
  )
  + theme_minimal()
  + theme(
    legend.position = "right",
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
)
```



### Shape difference of the caudal fin

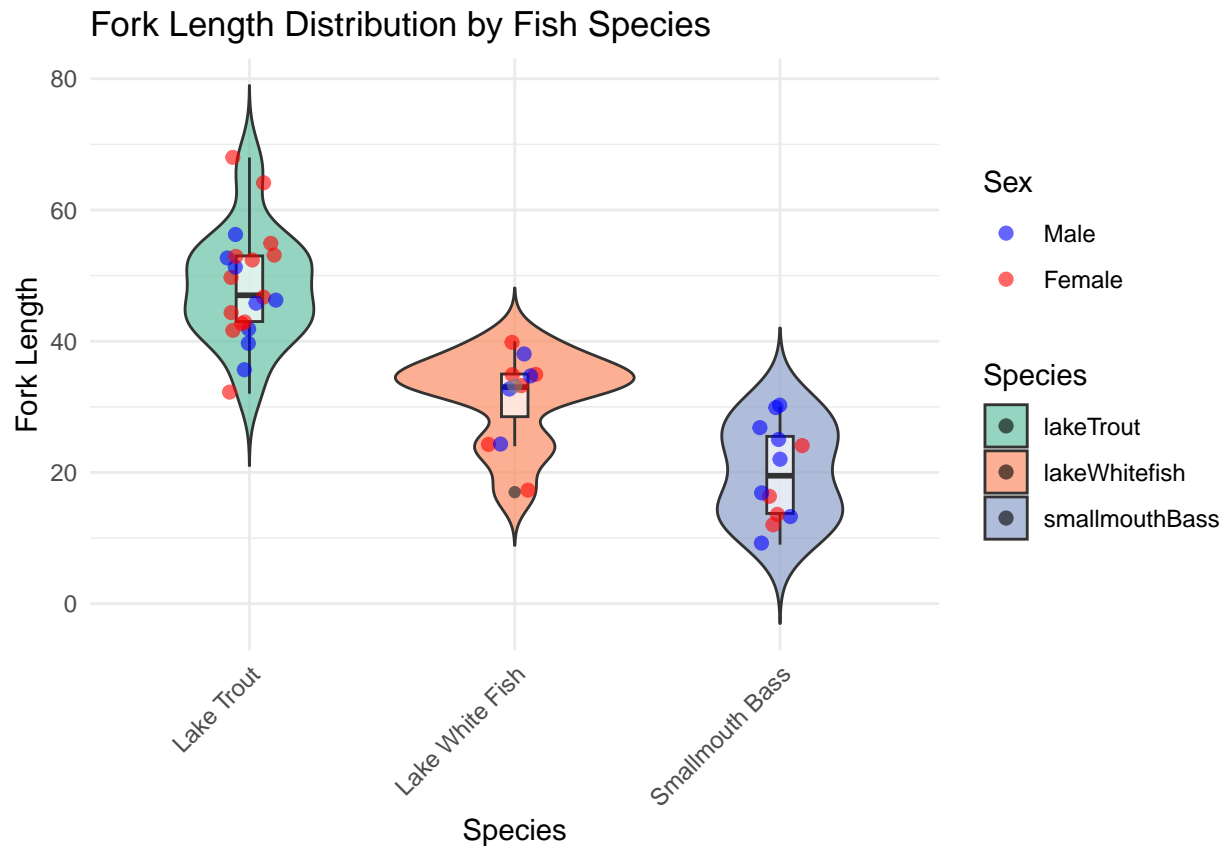
We want to use `total_length - fork_length` to compute the outermost portion of the caudal fin beyond its fork. We want to see if there's an obvious difference in the shape and the size of the caudal fin.

```
character_data$caudal_fin_diff <- character_data$totalLength - character_data$forkLength
(
  ggplot(character_data, aes(x = species, y = caudal_fin_diff, fill = species))
  + geom_violin(alpha = 0.7, trim = FALSE)
  + geom_boxplot(width = 0.1, fill = "white", alpha = 0.7)
  + geom_jitter(aes(color = factor(sex)), width = 0.1, alpha = 0.6, size = 2)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
                        labels = c("1" = "Male", "2" = "Female"),
                        name = "Sex")
  + scale_fill_brewer(palette = "Set2")
  + scale_x_discrete(labels = c("lakeTrout" = "Lake Trout",
                                "lakeWhitefish" = "Lake White Fish",
                                "smallmouthBass" = "Smallmouth Bass"))
  + labs(
    title = "Fork Length Distribution by Fish Species",
    x = "Species",
    y = "Fork Length",
    fill = "Species"
  )
  + theme_minimal()
  + theme(
```

```

legend.position = "right",
axis.text.x = element_text(angle = 45, hjust = 1)
)
)

```



### Weight-Length Curves

```

total_length_lm <- lm(log(weight) ~ log(totalLength), data = character_data)
summary(total_length_lm)

```

```

##
## Call:
## lm(formula = log(weight) ~ log(totalLength), data = character_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41943 -0.19619 -0.05407  0.23540  0.50269
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -10.9681     0.7551  -14.53  <2e-16 ***
## log(totalLength)  2.9142     0.1252   23.27  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2631 on 42 degrees of freedom

```

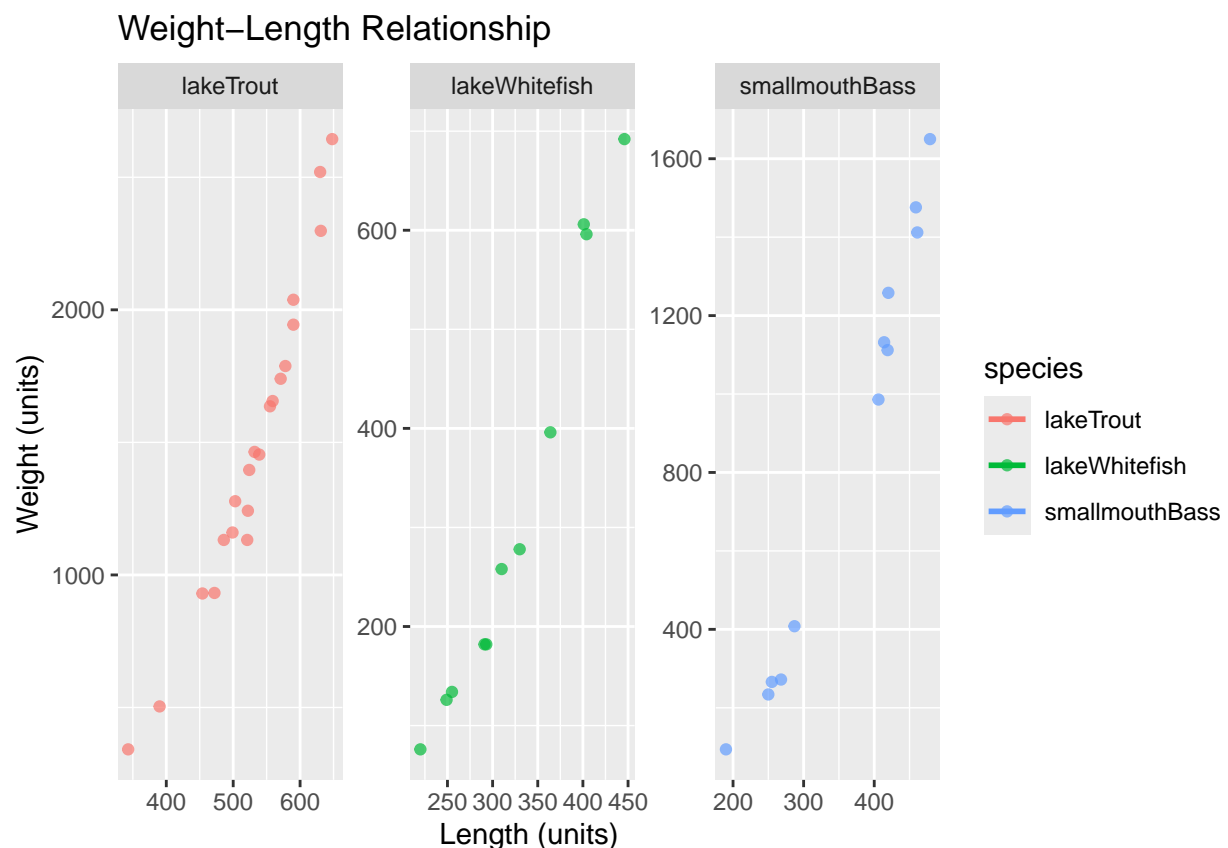
```
## Multiple R-squared:  0.928, Adjusted R-squared:  0.9263
## F-statistic: 541.6 on 1 and 42 DF,  p-value: < 2.2e-16

total_length_a <- exp(coef(total_length_lm)[1])
total_length_b <- coef(total_length_lm)[2]
ggplot(character_data, aes(x = totalLength, y = weight, color = species)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "lm", formula = y ~ total_length_a * I(x^total_length_b), se = FALSE) +
  labs(title = "Weight-Length Relationship",
       x = "Length (units)",
       y = "Weight (units)") +
  facet_wrap(~ species, scales = "free")

## Warning: Failed to fit group 1.
## Caused by error in `model.frame.default()`:
## ! variable lengths differ (found for 'total_length_a')

## Warning: Failed to fit group 2.
## Caused by error in `model.frame.default()`:
## ! variable lengths differ (found for 'total_length_a')

## Warning: Failed to fit group 3.
## Caused by error in `model.frame.default()`:
## ! variable lengths differ (found for 'total_length_a')
```



```
## linear regression models for weight-length parameters
wl_models <- by(character_data, character_data$species,
  function(data) {
```

```

        model <- lm(log(weight) ~ log(totalLength), data = data)
        a <- exp(coef(model)[1])
        b <- coef(model)[2]
        return(list(a = a, b = b))
    })

## parameters for each species
species_list <- levels(factor(character_data$species))
total_length_a <- sapply(species_list,
                        function(sp){
                            wl_models[[sp]]$a
                        })
total_length_b <- sapply(species_list,
                        function(sp){
                            wl_models[[sp]]$b
                        })
names(total_length_a) <- species_list
names(total_length_b) <- species_list

## predict standard w-l curves
predict_res <- data.frame()
## for loop to compute
for (sp in species_list) {
    ## find length range for each species
    length_range <- seq(
        min(character_data$totalLength[character_data$species == sp]),
        max(character_data$totalLength[character_data$species == sp]),
        length.out = 100
    )

    ## create standard weight for each species
    temp <- data.frame(
        totalLength = length_range,
        weight = total_length_a[sp] * length_range^total_length_b[sp],
        species = sp
    )

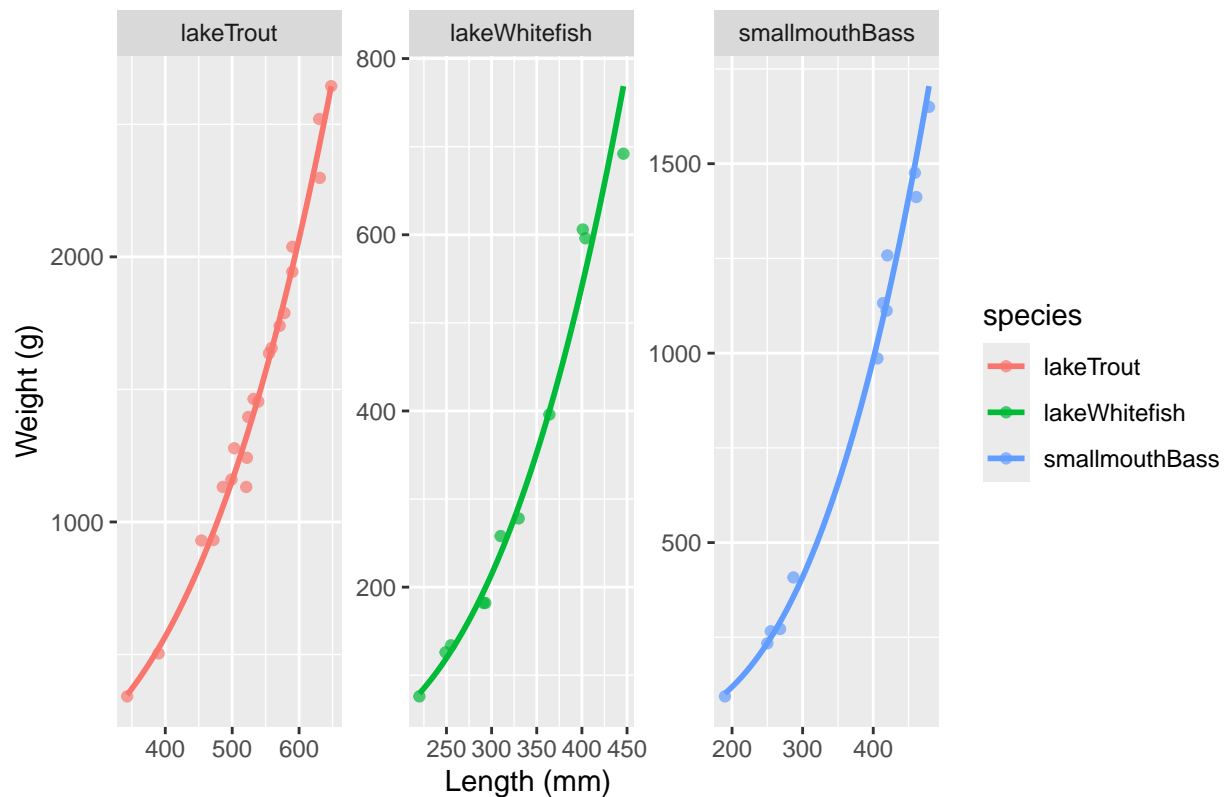
    predict_res <- rbind(predict_res, temp)
}

## plot
(
    ggplot(character_data, aes(x = totalLength, y = weight, color = species))
    + geom_point(alpha = 0.7)
    + geom_line(data = predict_res, aes(x = totalLength, y = weight, color = species), linewidth = 1)
    + labs(title = "Weight-Length Relationship",
           x = "Length (mm)",
           y = "Weight (g)")
    + facet_wrap(~ species, scales = "free")
)

```



## Weight–Length Relationship

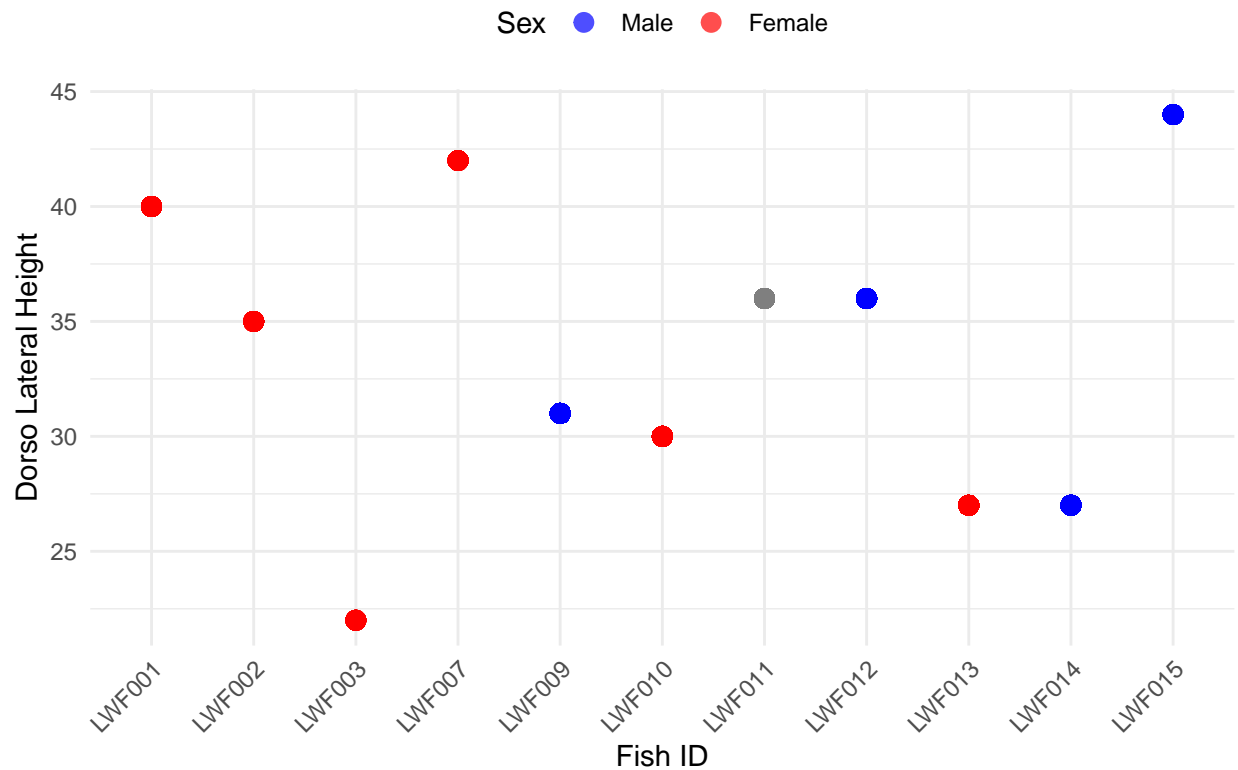


## Dorso Lateral Height

Individual fish dorso lateral height in each species.

```
## LakeTrout, LakeWhiteFish, SmallmouthBass
length_data = LakeWhiteFish
(
  ggplot(length_data, aes(x = fishNum, y = dorsoLatHeight))
  + geom_point(aes(color = sex), size = 3, alpha = 0.7)
  + geom_line(aes(group = fishNum), alpha = 0.3)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
                        labels = c("1" = "Male", "2" = "Female"),
                        name = "Sex")
  + labs(
    title = paste0("Indiviaul Dorso Lateral Height distribution in ", length_data$species),
    x = "Fish ID",
    y = "Dorso Lateral Height"
  )
  + theme_minimal()
  + theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "top"
  )
)
```

## Indivaiul Dorso Lateral Height distribution in lakeWhitefish



Dorso lateral height distribution comparison between three species.

```
(
  ggplot(character_data, aes(x = species, y = dorsoLatHeight, fill = species))
  + geom_violin(alpha = 0.7, trim = FALSE)
  + geom_boxplot(width = 0.1, fill = "white", alpha = 0.7)
  + geom_jitter(aes(color = factor(sex)), width = 0.1, alpha = 0.6, size = 2)
  + scale_color_manual(values = c("1" = "blue", "2" = "red"),
    labels = c("1" = "Male", "2" = "Female"),
    name = "Sex")
  + scale_fill_brewer(palette = "Set2")
  + scale_x_discrete(labels = c("lakeTrout" = "Lake Trout",
    "lakeWhitefish" = "Lake White Fish",
    "smallmouthBass" = "Smallmouth Bass"))
  + labs(
    title = "Dorso Lateral Height Distribution by Fish Species",
    x = "Species",
    y = "Dorso Lateral Height",
    fill = "Species"
  )
  + theme_minimal()
  + theme(
    legend.position = "right",
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
)
```

