Manual Approaches to the Unending Task of Labeling

aka... Sam and Kacey's Wild
Adventures in the Math Help Room

Advisors - Dr. Chris Tralie, Dr. Leslie New

Student Researchers – Sam Gregory, Kacey La

what we did

- we labeled whales

the end

1: TailLeft: [2468, 198]
2: TailRight: [1785, 41]
3: UpTailLeft: [2000, 486]
4: UpTailRight: [1942, 380]
5: BackLeft: [1946, 992]
6: BackRight: [1598, 958]
7: HeadLeft: [1740, 1473]
8: HeadRight: [1452, 1306]
9: Head: [1494, 1621]

[1785, 41]
[2468, 198]
[1942, 380]
[2000, 486]
[1598, 958]
[1946, 992]
[1452, 1306]
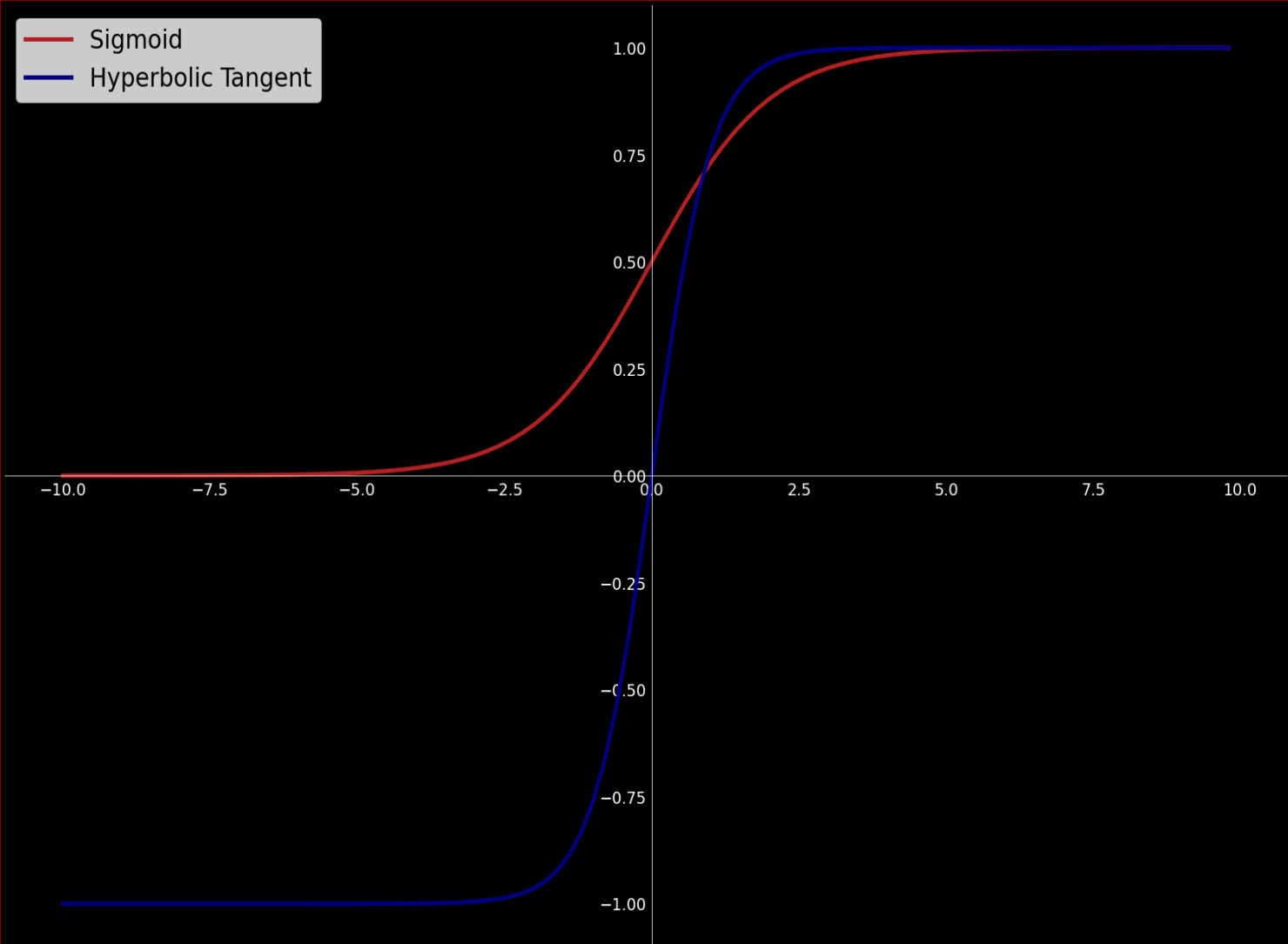[1740, 1473]
[1494, 1621]

Stage 1: Labeling

- We made some more labels. And then some more. And more.

Stage 2: Experimentation

$$S(x) = \frac{1}{1 + e^{-x}} \text{ vs. } H(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- Rotation

- Limit Training Data

- Activation Function
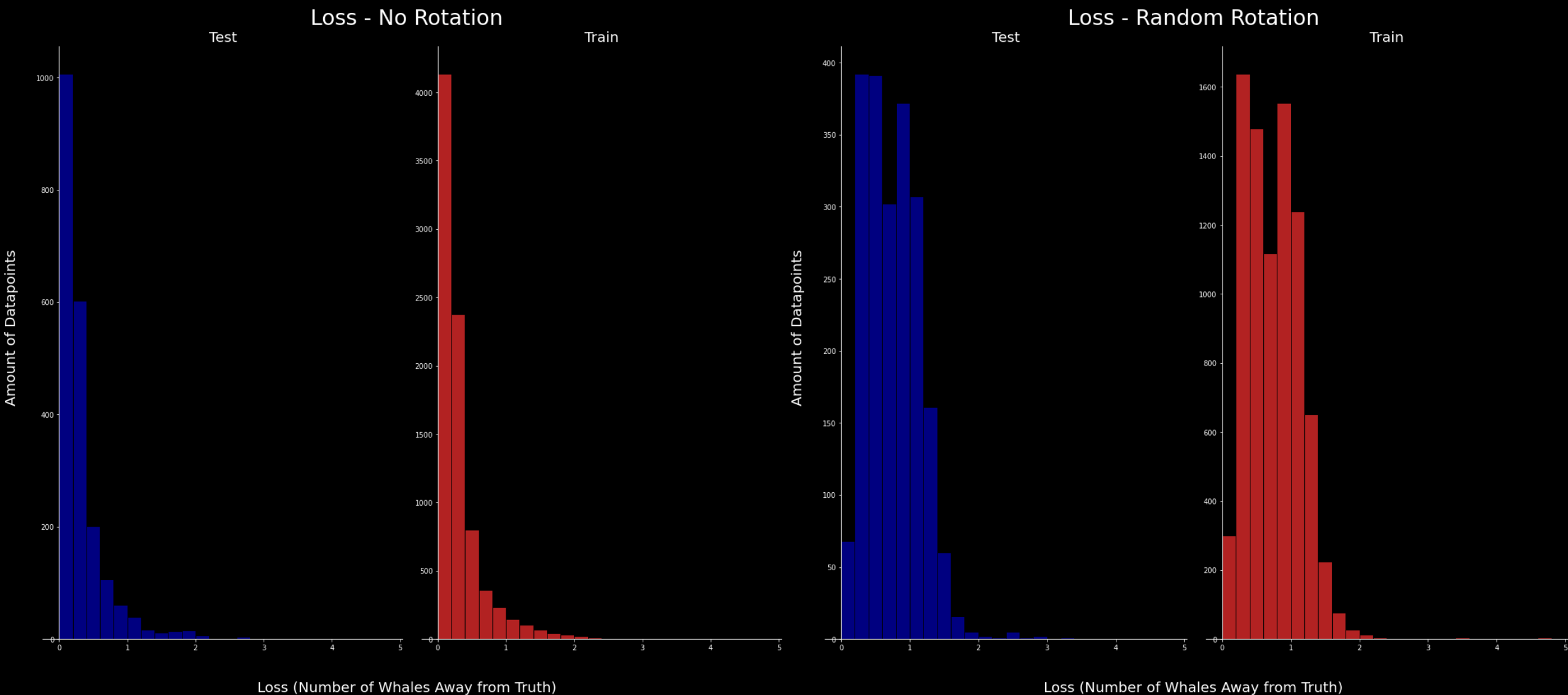 - Hyperbolic Tangent vs Sigmoid

- Number of Epochs


```
..........Epoch 118, Loss = 0.302072, Elapsed Time 305.688
..........Epoch 119, Loss = 0.301691, Elapsed Time 304.789
..........Epoch 120, Loss = 0.290739, Elapsed Time 300.408
..........Epoch 121, Loss = 0.293249, Elapsed Time 306.503
..........Epoch 122, Loss = 0.294827, Elapsed Time 303.068
..........Epoch 123, Loss = 0.282486, Elapsed Time 302.922
..........Epoch 124, Loss = 0.296332, Elapsed Time 303.023
..........Epoch 125, Loss = 0.286881, Elapsed Time 301.648
..........Epoch 126, Loss = 0.288068, Elapsed Time 303.341
..........Epoch 127, Loss = 0.295508, Elapsed Time 302.441
..........Epoch 128, Loss = 0.293827, Elapsed Time 303.195
..........Epoch 129, Loss = 0.272579, Elapsed Time 299.098
..........Epoch 130, Loss = 0.278180, Elapsed Time 302.141
..........Epoch 131, Loss = 0.278964, Elapsed Time 304.264
..........Epoch 132, Loss = 0.276265, Elapsed Time 307.606
..........Epoch 133, Loss = 0.287676, Elapsed Time 307.561
..........Epoch 134, Loss = 0.282605, Elapsed Time 304.529
..........Epoch 135, Loss = 0.270754, Elapsed Time 304.367
```
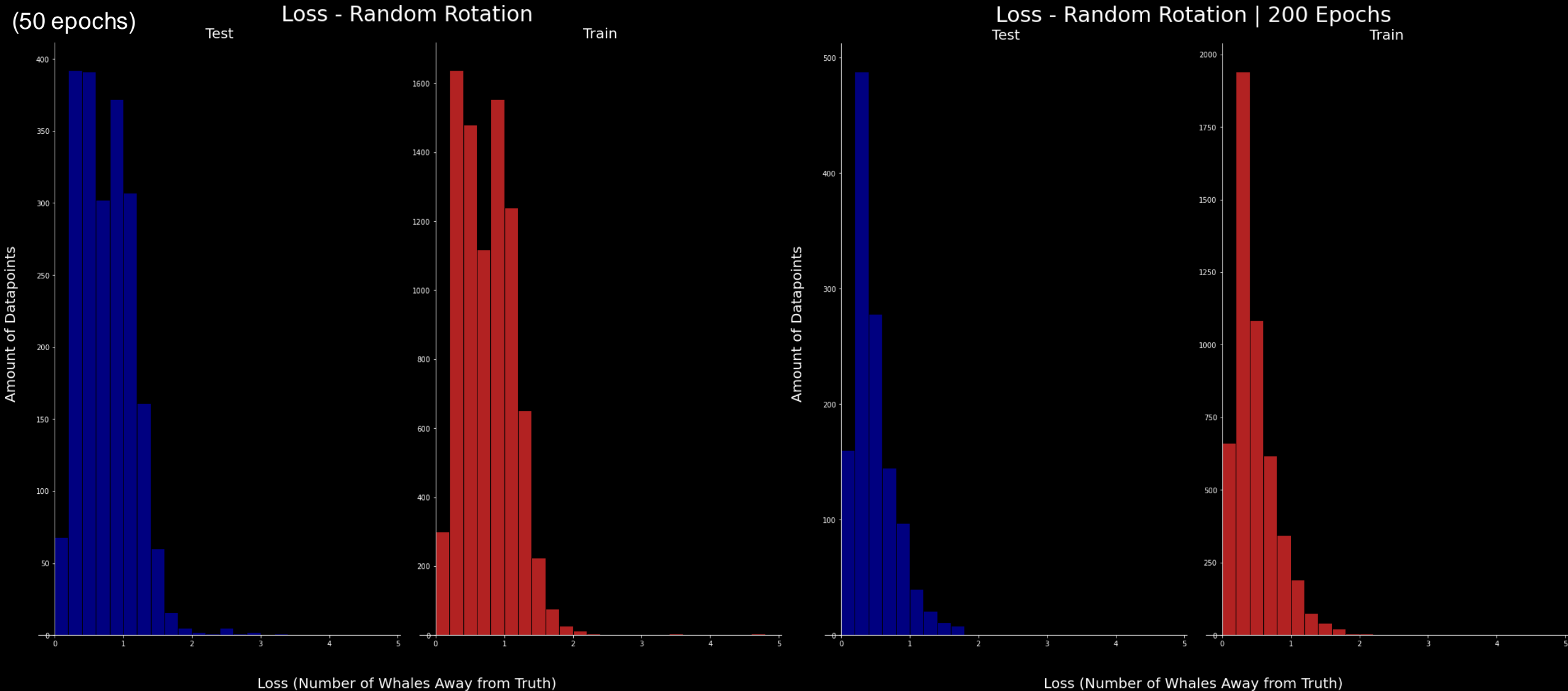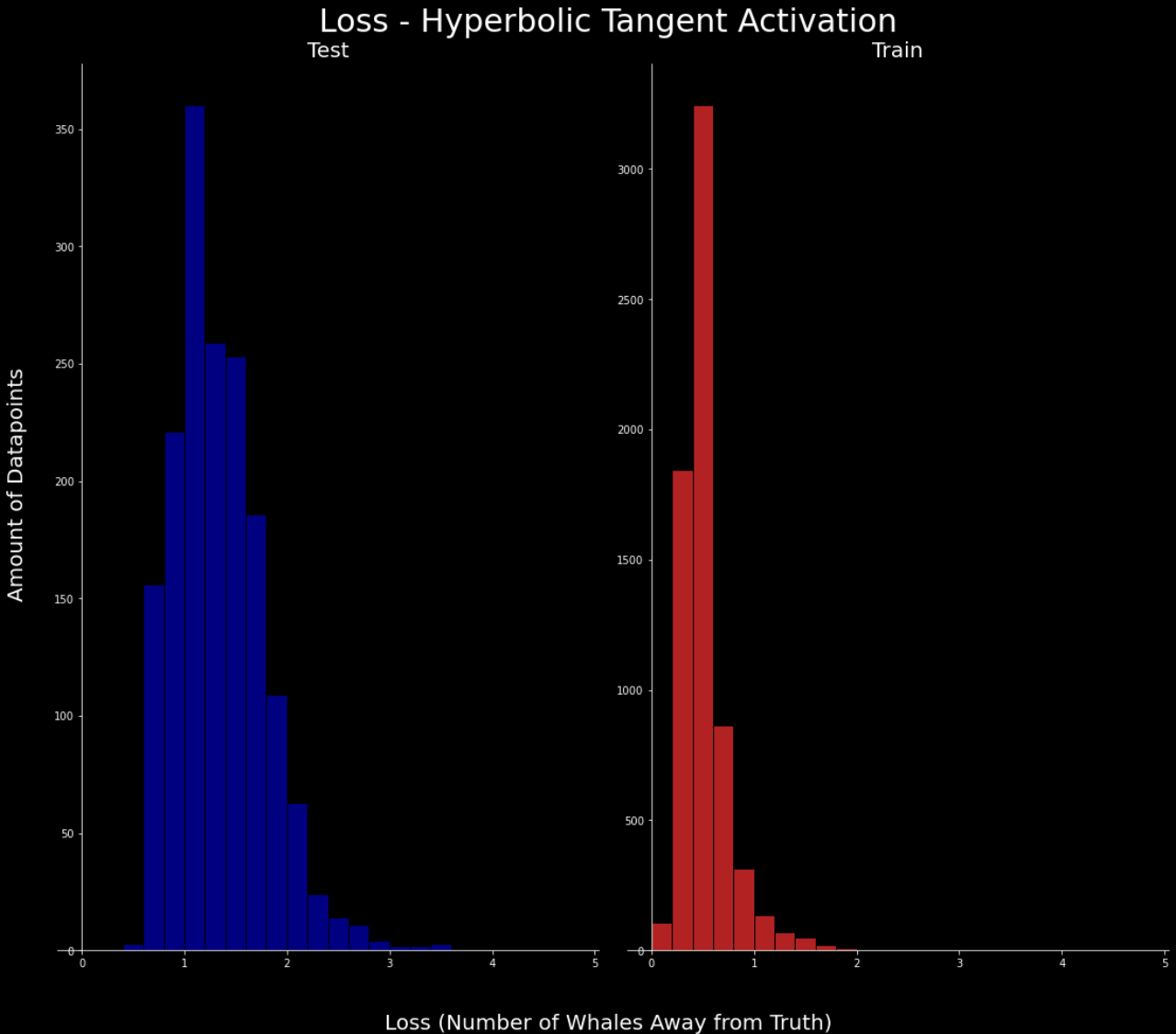
Stage 3: Spot-Check

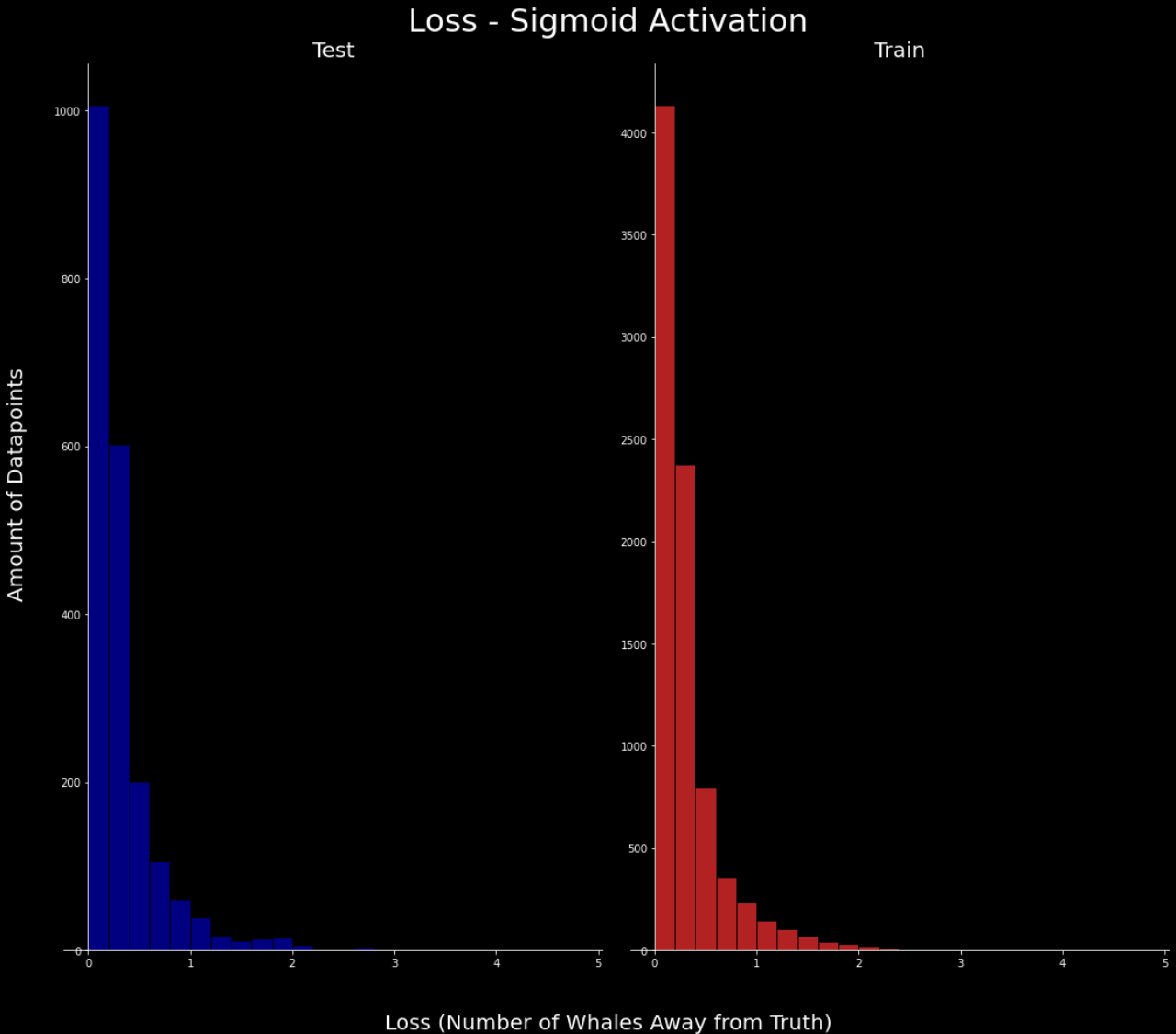# Stage 4: Histogram-ization

Loss - No Rotation

Test

Train

Loss - Random Rotation

Test

Train

Amount of Datapoints

Amount of Datapoints

Loss (Number of Whales Away from Truth)

Loss (Number of Whales Away from Truth)

Stage 4: Histogram-ization

Stage 4: Histogram-ization

Loss - Sigmoid Activation

Test   Train

Loss - Hyperbolic Tangent Activation

Test   Train

Amount of Datapoints

Loss (Number of Whales Away from Truth)
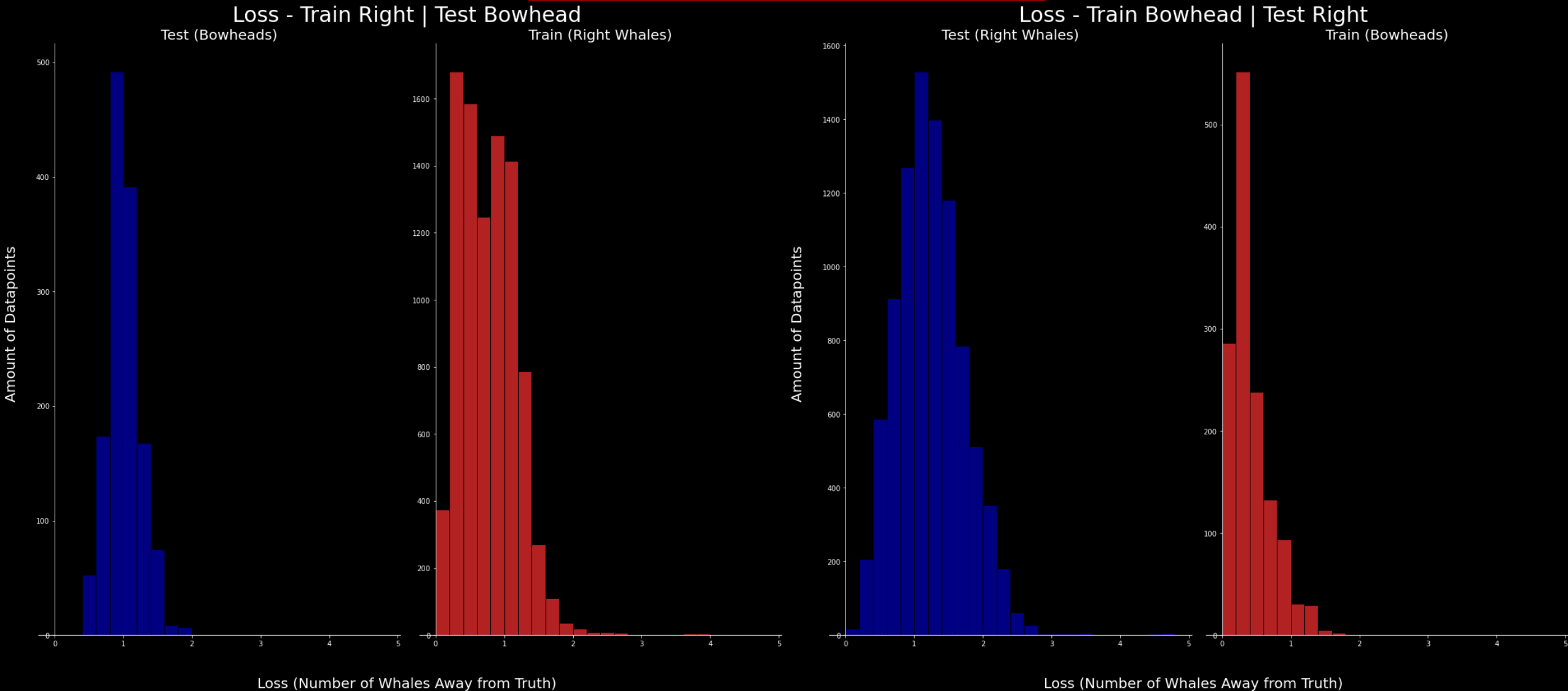
Amount of Datapoints

Loss (Number of Whales Away from Truth)

Stage 4: Histogram-ization

**both networks with random rotation

Stage 5: Grief

- Over 2,000 whales | 3 points per whale | ~3 whales per minute
  - Do it 3 times over
- ~34 Hours Spent *just labeling whales*
- Accounting for human error, review each whale to make the points consistent
- Over 2,000 whales | ~100 whales reviewed per hour
  - ~20 Hours Spent
  - How many times did Kacey label wrong side?

Stage 6: Acceptance

- Labeling isn't that bad.

- "The code works, I'm 100% confident.

  - ... unless it doesn't." – Kacey La

- "And True."

  - `while do_loop and True:` – Sam Gregory

Our Inspirational Leader