Palmer Penguins Part 1 Video: Complete Open-Source Production Guide

Building "EDA and Simple Regression" using LibreOffice Impress, Audacity, OBS Studio, Kdenlive, and Whisper

---

0. AVAILABLE ANALYSIS IMAGES

The following PNG files from the R analysis are available in the `posts/palmer_penguins_part1/` directory and should be used in your slides:

Filename	Description	Recommended Slide
`penguin-hero-part1.png`	Hero/title image	Slide 1: Title
`species-distribution.png`	Bar chart of species counts	Slide 6: Species Distribution
`island-distribution.png`	Distribution across islands	Slide 4: Dataset Overview (optional)
`morphometric-distributions.png`	Histograms of measurements	Slide 5: Data Exploration
`flipper-body-mass-relationship.png`	Main scatter plot (flipper vs mass)	Slides 7-8: Main Question & Patterns
`correlation-matrix.png`	Correlation heatmap	Slide 9: Correlation Coefficient
`eda-overview.png`	Combined species + relationship plot	Slide 4: Dataset Overview
`species-comparison.png`	Body mass boxplot by species	Slide 6: Species Distribution (alt)
`simple-regression-model.png`	Regression line with confidence band	Slide 13: Fitted Regression Line
`model-diagnostics.png`	Residual diagnostic plot	Slide 15: Model Limitations

Usage Notes: - All images are 300 DPI, suitable for 1920×1080 video - Images use the high-contrast color palette: - Adelie: #FF6B6B (coral) - Chinstrap: #9B59B6 (purple) - Gentoo: #2E86AB (deep ocean blue) - Import directly into LibreOffice Impress or Kdenlive timeline

---

1. LESSON STRUCTURE FOR PART 1

Video Title

"Palmer Penguins Part 1: Exploratory Data Analysis & Simple Regression"

SHORT VERSION (3 minutes) — RECOMMENDED

Duration: 3 minutes Instructional Arc:

1. **Hook & Data Overview** (30 sec)
   - Quick intro + 333 penguins, 3 species
   - Key question: Can flipper length predict body mass?
2. **Visual Exploration** (45 sec)
   - Species distribution
   - Scatter plot showing relationship
   - Correlation: r = 0.87
3. **Simple Regression** (1 min)
   - Model equation and key results

- $R^2 = 0.762$ (76% variance explained)
- Example prediction

4. **Limitations & Next Steps** (45 sec)
   - Species clustering in residuals
   - Preview Part 2: Adding species improves R^2 to 0.86

LONG VERSION (14–16 minutes) — OPTIONAL

Instructional Arc:

1. **Hook & Context** (1 min)
   - "Why do we care about penguin data?"
   - Real-world relevance: Antarctic ecology, conservation
2. **Meet the Data** (2.5 min)
   - Dataset overview
   - 333 penguins, 3 species
   - Key measurements: flipper length, body mass
3. **Exploratory Analysis** (4 min)
   - Species distribution
   - Relationship visualization
   - Correlation coefficient
4. **Simple Regression Model** (5 min)
   - Building the model in R
   - Interpreting coefficients
   - Model fit (R^2)
5. **Key Takeaways & Preview** (1.5 min)
   - What we learned
   - Why this matters
   - Teaser for Part 2

---

2. COMPLETE SCRIPT WITH TIMINGS

SHORT VERSION SCRIPT (3 minutes, 8 slides)

[SLIDE 1: Title]   Duration: 10 seconds IMAGE: penguin-hero-part1.png

Narration (25 words): "Palmer Penguins Part 1: Exploratory Data Analysis and Simple Regression. Can we predict a penguin's body mass just from its flipper length? Let's find out."

[SLIDE 2: The Data]   Duration: 20 seconds IMAGE: eda-overview.png

Narration (50 words): "We have 333 penguins from three species: Adelie, Chinstrap, and Gentoo. The scatter plot on the right already hints at something interesting—longer flippers seem to go with heavier bodies. Gentoo penguins, in blue, are clearly the largest. But how strong is this relationship?"

[SLIDE 3: Correlation]   Duration: 20 seconds IMAGE: correlation-matrix.png

Narration (45 words): "The correlation between flipper length and body mass is 0.87—that's strong. But correlation only tells us variables are related. To make predictions, we need regression. Let's fit a line through these points and see how well it works."

[SLIDE 4: The Model]   Duration: 25 seconds IMAGE: simple-regression-model.png

Narration (60 words): "Here's our simple linear regression. The equation: Body Mass equals negative 5,781 plus 49.7 times Flipper Length. That slope means every additional millimeter of flipper adds about 50 grams of body mass. The R-squared is 0.762—flipper length alone explains 76 percent of the variation in body mass. That's surprisingly good for one predictor."

[SLIDE 5: Making Predictions]   Duration: 20 seconds IMAGE: simple-regression-model.png (same, with callout overlay)

Narration (45 words): "What does this mean in practice? A penguin with 200-millimeter flippers? We predict about 4,100 grams. The gray band shows our uncertainty. Field researchers could use this to estimate body mass quickly—just measure the flipper."

[SLIDE 6: The Problem]   Duration: 25 seconds IMAGE: model-diagnostics.png

Narration (55 words): "But there's a problem. Look at the residuals—the prediction errors—colored by species. See how they cluster? Gentoo residuals sit together, separate from Adelie and Chinstrap. Our model treats all penguins the same, but species clearly matters. We're missing something important."

[SLIDE 7: Key Takeaways]   Duration: 15 seconds IMAGE: species-comparison.png

Narration (35 words): "What did we learn? Flipper length predicts body mass well—76 percent of variance. But species differences create patterns our simple model misses. The residuals tell us we can do better."

[SLIDE 8: Next Up]   Duration: 15 seconds IMAGE: penguin-hero-part1.png (or series graphic)

Narration (35 words): "In Part 2, we add species to the model. Spoiler: R-squared jumps from 0.76 to over 0.86. That's the power of thinking carefully about your data. See you next time."

Total: ~350 words   2:50 at natural pace

LONG VERSION SCRIPT (14 minutes, 18 slides)

[SLIDE 1: Title Slide]   Duration: 8 seconds

Narration (60 words): "Welcome to Palmer Penguins Part 1: Exploratory Data Analysis and Simple Regression. I'm [Your Name], and in this series, we're going to explore a real dataset collected by Antarctic researchers. We'll learn how data scientists approach a new dataset, build our first statistical model, and discover surprising relationships in penguin morphometrics. Let's dive in."

[SLIDE 2: Why Penguins?]

Duration: 12 seconds

Narration (95 words): "You might wonder: why penguins? Because real data is messy, interesting, and teaches us important lessons. The Palmer Penguins dataset was collected by Dr. Kristen Gorman at Palmer Station Antarctica between 2007 and 2009. It contains measurements of three penguin species: Adelie, Chinstrap, and Gentoo. Biologists use measurements like flipper length and body mass to understand penguin health. In conservation work, these measurements help identify if populations are thriving or struggling. And for us, it's a perfect playground for learning statistics."

[SLIDE 3: What We'll Learn]

Duration: 8 seconds

Narration (65 words): "By the end of this video, you'll understand three key concepts: First, how to explore a new dataset. Second, how to visualize relationships between variables. And third, how to build and interpret a simple linear regression model. These skills form the foundation for more advanced analysis. So let's start with the basics: getting to know our data."

[SLIDE 4: Dataset Overview]

Duration: 10 seconds

Narration (80 words): "The Palmer Penguins dataset contains 333 observations across three species. We have measurements including flipper length in millimeters, body mass in grams, bill length, and more. Here you see the sample size for each species: 146 Adelie penguins in coral, 68 Chinstrap penguins in purple, and 119 Gentoo penguins in deep blue. These color codes will help us track each species throughout our analysis."

[SLIDE 5: First Exploration - Code Segment]

Duration: 15 seconds

Narration (100 words): "When we load the data and examine it, we start with basic questions: How many observations do we have? Are there missing values? What's the structure of our data? Here we see a data glimpse showing our columns: species, flipper length, body mass, bill length, and others. We have 333 complete observations after removing rows with missing values. Notice how each species has different numbers of penguins. This imbalance is real data—no cleaning it away."

[SHOW: Console output of glimpse() and missing value counts]

[SLIDE 6: Species Distribution]

Duration: 12 seconds

Narration (85 words): "Let's visualize the composition. This bar chart shows our three species and their percentages. Adelie makes up 43.8 percent, Chinstrap 20.4 percent, and Gentoo 35.7 percent. We're also seeing the first interesting pattern: the species differ in average body mass. Adelie and Chinstrap are lighter, averaging around 3700 grams, while Gentoo penguins are noticeably heavier at 5076 grams. This difference will matter when we build our model."

[SHOW: Bar chart with percentages labeled]

[SLIDE 7: The Main Question]

Duration: 8 seconds

Narration (65 words): "Now the core question: can we predict body mass from flipper length? This is a practical question for field researchers. If you can quickly measure flipper length, can you estimate body weight? The scatter plot shows all 333 penguins, colored by species. Each point's horizontal position is its flipper length in millimeters. Its vertical position is body mass in grams."

[SHOW: Scatter plot]

[SLIDE 8: Visible Patterns]

Duration: 10 seconds

Narration (80 words): "Look closely at this scatter plot. You can see a strong positive relationship: longer flippers associate with heavier bodies. But notice something else: the three species form distinct clusters. Gentoo penguins, in deep blue, are separated from the other two. This clustering suggests that flipper length alone might not tell the whole story—species matters. We'll explore that in Part 2. For now, let's quantify this relationship with correlation."

[SLIDE 9: Correlation Coefficient]

Duration: 8 seconds

Narration (70 words): "The correlation between flipper length and body mass is 0.87. That's a strong positive correlation—close to one means they move together. But remember: correlation doesn't tell us the exact relationship. A correlation of 0.87 means the variables are strongly related, but doesn't yet give us a way to make predictions. For that, we need regression."

[SHOW: Calculated correlation value]

[SLIDE 10: Introducing Simple Linear Regression]

Duration: 10 seconds

Narration (75 words): "Simple linear regression finds the line that best fits our data points. We're fitting one predictor—flipper length—to predict one outcome: body mass. The model estimates two parameters: an intercept and a slope. The intercept is the predicted body mass when flipper length is zero. The slope tells us how much body mass increases for each additional millimeter of flipper length."

[SLIDE 11: Model Results - Code Segment]

Duration: 15 seconds

Narration (105 words): "Here's our fitted model. The intercept is negative 5781 grams. That doesn't make biological sense—we can't have negative body mass. But that's okay. The intercept exists in the mathematical space outside our data range. The slope is what matters: 49.7 grams per millimeter. This means for every one millimeter increase in flipper length, we predict an increase of 49.7 grams in body mass.

The R-squared is 0.762, meaning flipper length explains 76.2 percent of the variation in body mass. That's strong. The RMSE of 393 grams tells us our typical prediction error."

[SHOW: Regression output coefficients, R², RMSE]

[SLIDE 12: Model Equation]

Duration: 8 seconds

Narration (60 words): "We can write our model as a simple equation: Body Mass equals negative 5781 plus 49.7 times Flipper Length. In statistics notation, we write y-hat—the predicted value—equals our intercept plus slope times x. This is our predictive model. Given a flipper length, we can predict body mass."

[SHOW: Mathematical equation]

[SLIDE 13: Visualizing the Regression]

Duration: 12 seconds

Narration (85 words): "This plot shows the scatter of data with the fitted regression line in black. The gray band around the line is our 95 percent confidence interval—our uncertainty about where the true line lies. Notice the line runs through the middle of the cloud of points. Some points sit above, some below. Those deviations are our residuals. The bigger the deviations, the worse our fit. In this case, the fit is pretty good, but not perfect."

[SHOW: Scatter plot with regression line and confidence band]

[SLIDE 14: Example Predictions]

Duration: 10 seconds

Narration (70 words): "Let's make some predictions. For a penguin with 180-millimeter flippers, we predict 3,082 grams. For 200 millimeters, 4,079 grams. For 220 millimeters, 5,075 grams. These predictions come with confidence intervals shown in brackets. The 95 percent confidence interval tells us where we're 95 percent confident the true line passes. Predictions are most precise in the middle of our data range, where we have more information."

[SHOW: Prediction table with confidence intervals]

[SLIDE 15: Model Limitations]

Duration: 12 seconds

Narration (90 words): "Before we celebrate, let's be honest about limitations. First, species matters. Our residuals cluster by species. Second, this model ignores the fact that Gentoo penguins are fundamentally different from Adelie and Chinstrap. Third, we're extrapolating carefully. We have flipper measurements between 172 and 231 millimeters. Predictions outside that range are unreliable. Finally, this data spans 2007 to 2009. Climate change may have affected these relationships. Good science acknowledges what we don't know."

[SHOW: Residual diagnostic plot with species clustering]

[SLIDE 16: Why This Matters]

Duration: 10 seconds

Narration (80 words): "For field researchers in Antarctica, this model is practical. It lets them estimate body mass quickly without scales. For conservation, tracking body mass trends helps detect population stress. For us, this simple model shows the power of exploratory data analysis and regression. It raises the right questions: what's missing? How do species differ? Can we build a better model? These are the questions we'll answer in Part 2."

[SLIDE 17: Key Takeaways]

Duration: 10 seconds

Narration (75 words): "Here's what we've learned: One, exploratory data analysis reveals patterns in raw data. Two, flipper length strongly predicts body mass, explaining 76 percent of variation. Three, linear regression gives us both predictions and understanding. And four, acknowledging limitations makes our science more credible. In Part 2, we'll add species to our model and watch our R-squared jump from 0.76 to over 0.86."

[SLIDE 18: Next Episode Preview]

Duration: 8 seconds

Narration (65 words): "What happens when we account for species differences? Spoiler: the model improves dramatically. We'll add a categorical predictor, learn about interaction effects, and see how good statistical thinking builds on simple foundations. Subscribe or check back next week for Part 2: Multiple Regression and Species Effects. Thanks for learning with us."

Total Script Word Count: ~1,540 words   14 minutes at natural speaking pace

---

3. SLIDE DECK SPECIFICATIONS (LibreOffice Impress)

Design System

Dimensions: - Format: 16:9 widescreen - Resolution: 1920 × 1080 pixels

Color Palette (high-contrast for accessibility):

```
Adelie:    #FF6B6B (coral)
Chinstrap: #9B59B6 (purple)
Gentoo:    #2E86AB (deep ocean blue)
Primary Text: #2C3E50 (dark gray)
Background: #FFFFFF (white)
Accent: #27AE60 (green for highlights)
```

Typography: - **Headings:** Inter or Source Sans Pro, 56pt, bold, #2C3E50 - **Body text:** Inter or Source Sans Pro, 32pt, regular, #2C3E50 - **Code/terminal:** Courier New or Liberation Mono, 20pt, monospace - **Slide numbers:** 14pt, light gray, bottom right

Layout Rules: - Top margin: 60 pixels - Side margins: 80 pixels - Bottom margin: 60 pixels - One concept per slide - Maximum 2 bullet points or 1 image per content slide - Minimum 40% whitespace

---

SHORT VERSION: 8-Slide Deck

Slide	Image	Key Content
1. Title	`penguin-hero-part1.png`	Title + "Can flipper length predict body mass?"
2. The Data	`eda-overview.png`	333 penguins, 3 species, scatter plot preview
3. Correlation	`correlation-matrix.png`	r = 0.87, strong relationship
4. The Model	`simple-regression-model.png`	Equation + $R^2 = 0.762$
5. Predictions	`simple-regression-model.png`	200mm → 4,100g example
6. The Problem	`model-diagnostics.png`	Species clustering in residuals
7. Takeaways	`species-comparison.png`	76% explained, but species matters
8. Next Up	`penguin-hero-part1.png`	Part 2 preview: R^2 → 0.86

LONG VERSION: 18-Slide Deck (Slide-by-Slide Templates)

Slide 1: Title Slide

```
Background: #FFFFFF
Image: penguin-hero-part1.png (positioned left or as background overlay)
Center:
  Title: "Palmer Penguins Part 1"
  Subtitle: "Exploratory Data Analysis & Simple Regression"
  Bottom: "Your Name | UCSD Biostatistics"
  Footer: Light gray bar, dark gray text
```

Slide 2: Context/Why Penguins

```
Left: Image of Antarctic penguins (free stock image)
Right:
  Heading: "Why Penguins?"
  Bullet 1: "Real ecological data"
  Bullet 2: "3 species, 333 observations (2007-2009)"
  Bullet 3: "Practical conservation applications"
```

Slide 3: Learning Objectives

```
Background: Light accent bar at top
Heading: "What We'll Learn"
Numbered list (use penguin species colors for bullets):
  1. Explore a new dataset (coral #FF6B6B)
  2. Visualize relationships (purple #9B59B6)
  3. Build regression models (deep blue #2E86AB)
```

Slide 4: Dataset Overview

```
IMAGE: eda-overview.png (full width, 2/3 of slide height)
  - Shows species distribution bar chart AND flipper-body mass relationship
  - Combined visualization provides context for both sample sizes and key patterns

Below image: Brief text callout
  "333 penguins across 3 species | Strong flipper-mass relationship visible"
```

Slide 5: Data Exploration Code

```
Left (1/2 of slide): Dark background (#F5F5F5) code block:
  library(palmerpenguins)
  data(penguins)
```

```
penguins_clean <- penguins %>% drop_na()
glimpse(penguins_clean)
```

Right (1/2 of slide): IMAGE: morphometric-distributions.png
  - Shows distribution of key measurements
  - Visual context for the data structure being explored

Slide 6: Species Distribution

IMAGE: species-distribution.png (large, filling 2/3 of slide)
  - Bar chart with species counts and percentages
  - Colors: Adelie (#FF6B6B), Chinstrap (#9B59B6), Gentoo (#2E86AB)

Alternative: species-comparison.png (boxplot showing body mass by species)
  - Use if you want to emphasize size differences between species

Below chart:
  Callout box: "Adelie most common (44%), Chinstrap least (20%)"

Slide 7: The Central Question

Heading (large): "Can flipper length predict body mass?"
Subheading: "A practical question for field researchers"

IMAGE: flipper-body-mass-relationship.png (right half of slide, partial reveal)
  - Scatter plot showing the relationship
  - Faded or cropped to tease the pattern without full reveal

Left side: White space with centered question text

Slide 8: Scatter Plot with Visible Pattern

IMAGE: flipper-body-mass-relationship.png (full screen, 3/4 of slide)
  - Full scatter plot now revealed
  - X-axis: Flipper Length (mm)
  - Y-axis: Body Mass (g)
  - Points colored by species with alpha transparency

Callout box (bottom right, overlay on image):
  "Strong positive relationship"
  "Species clustering visible"

Slide 9: Correlation Coefficient

Left side:
  Heading: "Measuring Linear Relationship"
  Large highlighted box:
    r = 0.87
    "Strong positive correlation"
    (in color: #4ECDC4)

  Bullet points:
    • 0.87 indicates strong relationship
    • -1 to +1 scale
    • Still: doesn't give us predictions

Right side: IMAGE: correlation-matrix.png (1/3 of slide width)
  - Shows full correlation matrix for context

- Highlights flipper-body mass correlation (0.87)

Slide 10: Linear Regression Introduction

Heading: "Simple Linear Regression"
Equation centered, large font:
  y =   + x
  or
  Body Mass = Intercept + Slope × Flipper Length

Below: Bullet points
  •   (intercept): predicted value when x = 0
  •   (slope): change in y per unit of x
  • Fits a line to minimize distance to points

Slide 11: Regression Results

Code block (dark background):

lm(body_mass_g ~ flipper_length_mm, data = penguins_clean) Coefficients: (Intercept) –5781.4 flipper_length_mm 49.7

R-squared: 0.762 RMSE: 392.8 grams

Callout boxes:
  "49.7 grams per mm" (highlighted, color #4ECDC4)
  "Explains 76.2% of variance"

Slide 12: Model Equation

Large centered equation:
  ŷ = –5781.4 + 49.7 × Flipper Length

Below (smaller text):
  "ŷ (y-hat) = predicted value"
  "Given a flipper measurement, we can predict body mass"

Visual: Simple graphic showing x → box (equation) → y

Slide 13: Fitted Regression Line

IMAGE: simple-regression-model.png (large, 3/4 of slide)
  – Scatter plot with fitted regression line in black
  – 95% confidence band shown in gray
  – Points colored by species (alpha = 60%)
  – Clear axis labels with units

Callout (top right, overlay):
  "Gray band = 95% confidence interval"
  "Line fits well, but not perfectly"

Slide 14: Making Predictions

Heading: "Example Predictions (with 95% CI)"

Table (centered):
Flipper Length	Predicted Body Mass	95% Confidence Interval
180 mm	3,082 g	[2,904, 3,260]

```
200 mm | 4,079 g | [3,981, 4,177]
220 mm | 5,075 g | [4,876, 5,274]
```

Slide 15: Model Limitations

```
Left side (1/2):
  Heading: "Be Honest About Limitations"

  Four callout boxes (stacked):
      Species clustering - "Residuals cluster by species"
      Data range - "Only valid for 172-231 mm"
      Temporal scope - "Data from 2007-2009"
      Missing predictors - "Species and other variables matter"

Right side (1/2): IMAGE: model-diagnostics.png
  - Residual diagnostic plot showing species clustering
  - Visual evidence that species matters
```

Slide 16: Practical Applications

```
Heading: "Why This Matters"

Three boxes with icons:
    Field Research
     "Estimate mass without scales"

    Conservation
     "Track population health trends"

    Statistical Foundation
     "Simple models teach important concepts"
```

Slide 17: Key Takeaways

```
Background: Light blue accent bar at top

Heading: "In One Slide"

Numbered takeaways (penguin colored bullets):
  1. EDA reveals patterns (coral)
  2. Flipper → Body Mass is strong (teal)
  3. Regression gives predictions & understanding (light blue)
  4. Always acknowledge limitations
```

Slide 18: Next Episode

```
Background: Gradient light blue to white

Center:
  Large text: "Part 2: Multiple Regression"
  Subtitle: "Adding species improves R² from 0.76 to 0.86"

Bottom: Subscribe prompt / Next episode date
```

---

4. RECORDING NARRATION IN AUDACITY

Software: Audacity 3.7+ (manual.audacityteam.org)

Setup

1. **Open Audacity** (free download: audacityteam.org)
2. **Create a new project:** File → New
3. **Set recording to Mono** (via Audio Setup Toolbar - preferred method):
   - Click **Audio Setup** button in the toolbar
   - Select **Recording Channels → 1 (Mono)**
   - Or: Audacity → Settings (Cmd+,) → Audio Settings → Channels → 1
4. **Select microphone:**
   - Click **Audio Setup** button → **Recording Device**
   - Select your microphone from the list
5. **Test levels:**
   - Click microphone icon to enable monitoring
   - Speak normally—meter should peak between **-12 dB and -6 dB**

Recording Workflow

For each narration segment:

1. **Prepare:** Read the script 1–2 times silently to internalize pacing
2. **Place cursor:** Click in the audio track where you want to record
3. **Hit Record** (or Ctrl+R / Cmd+R)
4. **Speak clearly:** Natural pace, not too fast
   - Pause 1–2 seconds between major ideas
   - Breathe naturally (edit out long breaths later)
5. **Stop recording** (Spacebar or click stop button)

Processing Each Narration Track

After recording one segment:

1. **Normalize:**
   - Select all (Ctrl+A)
   - Effect → Normalize
   - Default settings (-1.0 dB)   OK
   - *This brings volume to consistent level*
2. **Noise Reduction:**
   - Select a small section of "silence" (room tone)
   - Effect → Noise Reduction
   - Click "Get Noise Profile"
   - Select all audio again (Ctrl+A)
   - Effect → Noise Reduction again
   - Amount: 12 dB   OK
   - *Removes background hum/hiss*
3. **Gentle Compression:**
   - Select all (Ctrl+A)
   - Effect → Compressor
   - Default settings (ratio 2:1, threshold –20 dB)   OK
   - *Smooths volume peaks*
4. **Fade In/Out:**
   - Select first 0.5 seconds of audio
   - Effect → Fade In
   - Select last 0.5 seconds
   - Effect → Fade Out
   - *Prevents clicks at start/end*

Exporting Narration

Export as WAV (highest quality for editing): - File → Export → Export as WAV - Naming: `slide01_narration.wav`, `slide02_narration.wav`, etc. - Quality: 16-bit PCM

Or **for final delivery**, export as MP3 at 192 kbps

Tips for Recording Quality

- **Microphone placement:** 6–8 inches from mouth, slightly off-axis
- **Pacing:** Speak at 120–140 words per minute (slower than normal)
- **Pauses:** Natural 1–2 second pauses between slides
- **Breath control:** Breathe between sentences, not mid-thought
- **One take rule:** Record entire slide narration without stopping
- **Multiple takes:** Do 2–3 takes, pick the best one
- **Studio quality:** Record in a small room (closet, car with blankets) for better acoustics

---

5. SCREEN CAPTURE WITH OBS STUDIO

Software: OBS Studio 32.x (obsproject.com/kb)

Setup

1. **Install OBS Studio** (free: obsproject.com)
2. **Create a new scene:**
   - Click "+" under Scenes
   - Name: "Palmer_Penguins_Part1"

Adding Sources

Source 1: LibreOffice Impress Presentation

1. Click "+" under Sources
2. Select "Display Capture" (or "Window Capture" for specific window)
3. Select your Impress window
4. Scale to fit (right-click source → Transform → Fit to Screen)

Optional Source 2: Webcam (if you want a picture-in-picture) - Click "+" under Sources - Select "Video Capture Device" → your webcam - Resize/position in corner (small window) - Optional: add rounded corners filter

Recording Settings

1. **Settings → Output → Recording**
   - Output Mode: **Advanced** (for more control)
   - Recording Path: `/path/to/your/project/` (create folder)
   - Recording Format: **MKV** (recoverable if OBS crashes; remux to MP4 later)
   - Video Encoder: NVENC (NVIDIA), AMF (AMD), or x264 (CPU)
   - Audio Encoder: AAC
   - Audio Bitrate: 192 kbps
     Tip: After recording, use File → Remux Recordings to convert MKV to MP4.
2. **Settings → Video**
   - Base Canvas: 1920×1080
   - Output Scaled: 1920×1080

- FPS: 30
3. **Settings → Audio**
   - Sample Rate: 48 kHz
   - Channels: Stereo

Recording Workflow

1. **Open your slide deck** (LibreOffice Impress in slideshow mode)
2. **Start the presentation**
3. **In OBS:** Click "Start Recording"
4. **Advance slides** at your own pace (you'll sync narration in editing)
5. **Finish slideshow**
6. **In OBS:** Click "Stop Recording"

Output file: `slide_capture_raw.mp4` in your project folder

5.5 AI AVATAR OPTION (INTRO/OUTRO HYBRID)

An alternative to webcam PiP: use AI-generated talking avatar for intro and outro segments, with pure voiceover for the main educational content.

Why This Approach Works

- **Personal connection:** Avatar intro/outro adds human element
- **Clean content:** No webcam distraction during statistical explanations
- **Professional polish:** Music bookends signal "beginning" and "end"
- **No equipment:** Works without webcam, ring light, or studio setup

Timeline Structure

```
+--------+----------------------------------------+--------+
| INTRO  |        SLIDES 3-17 (VOICEOVER ONLY)    | OUTRO  |
| Avatar |      No music - pure narration focus   | Avatar |
| +Music |                                        | +Music |
| ~20sec |                ~13 minutes             | ~20sec |
+--------+----------------------------------------+--------+
```

AI Avatar Platforms

Platform	Pricing	Best For
HeyGen	Free tier + paid	High quality, many avatar styles
Synthesia	Subscription	Enterprise, multilingual
Vidnoz	Free tier	Quick tests, budget-friendly
D-ID	Free tier + paid	Creative Studio, custom avatars

Workflow

1. **Prepare avatar scripts:**
   - **Intro** (Slides 1-2 narration, ~120 words): Opening hook and context
   - **Outro** (Slide 18 narration, ~65 words): Summary and preview
2. **Generate avatar clips:**
   - Sign up for chosen platform (HeyGen recommended)
   - Select avatar style (professional, academic)

- Paste intro script → Generate → Download MP4
- Paste outro script → Generate → Download MP4
3. **Record voiceover for slides 3-17:**
   - Use Audacity as described in Section 4
   - No avatar needed for educational content
4. **Assemble in Kdenlive:**
   - V1: Avatar intro → Slide images 3-17 → Avatar outro
   - A1: Avatar audio (built-in) → Voiceover WAVs → Avatar audio
   - A2: Music (intro, -15dB) → silence → Music (outro, -15dB)

Avatar Tips

- **Choose consistent style:** Same avatar throughout series
- **Background:** Simple, uncluttered (most platforms offer options)
- **Voice:** Match avatar voice to your voiceover style if possible
- **Length:** Keep avatar segments short (10-30 seconds each)

Alternative: Simpler Approach (Export Slides as Images)

If you want maximum control over timing:

1. **In LibreOffice Impress:** File → Export as PDF

2. **Convert PDF to individual PNG images** (using ImageMagick or online tool):

   ```
   convert -density 150 -quality 90 slides.pdf slides.png
   # Generates: slides-0.png, slides-1.png, slides-2.png, etc.
   ```

3. **Kdenlive will handle timing** (see next section)

---

6. EDITING VIDEO WITH KDENLIVE

Software: Kdenlive 25.x (docs.kdenlive.org)

Project Setup

1. **Open Kdenlive** (free: kdenlive.org)
2. **Create new project:**
   - File → New
   - FPS: 30
   - Resolution: 1920×1080 (16:9)
   - Audio: 48 kHz, stereo

Workflow: Building Your Timeline

Option A: Using screen captures + narration

1. **Import screen capture:**
   - Project → Add Clip
   - Select `slide_capture_raw.mp4`
   - Drag to Video Track 1
2. **Import narration audio files:**
   - Project → Add Clip
   - Select all `slide0X_narration.wav` files
   - Drag to Audio Track 1 below corresponding videos

3. **Sync narration to slides:**
   - Use timeline markers to match narration timing to slide transitions
   - Adjust video playback speed slightly if needed to fit timing

Option B: Using slide images + narration (more control)

1. **Import all PNG slide images:**
   - Project → Add Clip → select all PNG files
2. **Drag to timeline:**
   - Slide 1 image → Video Track 1
   - Slide 2 image → Video Track 1
   - (One per track or in sequence on same track)
3. **Set image durations:**
   - Right-click each image → Duration
   - Duration = length of narration for that slide
   - (e.g., 8-second slide = 8 seconds duration)
4. **Add narration:**
   - Drag `slide01_narration.wav` to Audio Track 1
   - Drag `slide02_narration.wav` to Audio Track 1 after first
   - (Timeline auto-arranges)
5. **Sync timing:**
   - Play timeline, watch if audio/video align
   - Trim/adjust individual clips as needed

Adding Transitions

1. **Select slide/clip** in timeline
2. **Right-click → Add Transition**
3. **Choose:** Dissolve (recommended, gentle)
4. **Duration:** 300–500 ms (0.3–0.5 seconds)
5. **Apply to all** (if consistent)

Transition tips:

- Use dissolve for most slide transitions
- Keep duration consistent throughout
- Avoid fancy transitions (wipes, spins) for educational content

---

6.5 POLISH EFFECTS (PROFESSIONAL QUALITY)

These techniques add professional polish without overwhelming educational content.

Ken Burns Effect (Subtle Zoom/Pan)

Adds visual interest to static analysis images.

In Kdenlive:

1. Select an image clip on timeline
2. Go to **Effects** tab → search "Transform"
3. Add **Transform** effect to clip
4. Set keyframes:
   - **Start:** Scale 100%, Position centered
   - **End:** Scale 105%, Position slightly shifted
5. Duration: Match slide duration (slow, subtle movement)

Recommended settings:

- Scale change: 100% → 105% (subtle zoom in)
- Position drift: 0-20 pixels over slide duration
- Apply to: Main analysis images (data visualizations)
- Skip for: Title slide, conclusion slide

Lower Third (Name Overlay)

Professional touch showing presenter name on title slide.

In Kdenlive:

1. **Create text clip:**
   - Project → Add Title Clip
   - Background: Semi-transparent black (#000000, 60% opacity)
   - Text: "Your Name | UCSD Biostatistics"
   - Font: Inter or Source Sans Pro, 28pt, white
   - Position: Bottom left, 80px from edges
2. **Add to timeline:**
   - Place on V2 track above slide-1
   - Duration: 5-8 seconds
3. **Animate (optional):**
   - Add Slide effect: animate in from left over 0.3 seconds
   - Fade out at end

Fade In/Out (Video Bookends)

Fade In (start of video):

1. Select first clip on timeline
2. Right-click → Add Effect → Fade → Fade In
3. Duration: 1 second

Fade Out (end of video):

1. Select last clip on timeline
2. Right-click → Add Effect → Fade → Fade Out
3. Duration: 1 second

Audio fades:

1. Select audio track
2. Drag the small triangles at start/end of audio clip
3. Create 0.5-second audio fades to prevent pops

Adding Background Music (Intro/Outro Only)

Important: Per cognitive load research, background music during statistical explanations hurts learning retention. Use music only for intro and outro segments—NOT during the main educational content.

Why intro/outro only:

- StatQuest, 3Blue1Brown use NO music during explanations
- Music competes with working memory during complex content
- Intro/outro music signals "beginning" and "end" professionally

1. **Download royalty-free track:**
   - Pixabay Music - Free, no attribution

- FreePD - Public domain
- YouTube Audio Library
- Find instrumental, minimal style (no lyrics)
2. **Import to Audio Track 2:**
   - Project → Add Clip → select music file
   - Drag to Audio Track 2
3. **Position music (intro/outro only):**
   - Place under title slide (intro): ~8-10 seconds
   - Place under final slide (outro): ~10-15 seconds
   - **Leave silence** under all educational content slides
4. **Adjust volume:**
   - Right-click music track → Volume
   - Set to **–15 to –18 dB** (much quieter than narration at –6 dB)
5. **Fade in/out:**
   - End of intro music: 1-second fade out
   - Start of outro music: 1-second fade in
   - Drag edges of music clip or use Fade effects

```
Audio Timeline Example (14-min video):
A1: [narration throughout all slides]
A2: [music]                                      [music]
    ↑ intro (fade out)               (fade in) ↑ outro
    slides 1-2                               slides 17-18
```

Adding Text Overlays (Optional but recommended)

For code sections and key statistics:

1. **Insert text clip:**
   - Project → Add Clip → Color Clip
   - Drag to Video Track 2 above slides
   - Duration: match your narration
2. **Add text:**
   - Right-click → Edit Clip Text
   - Type: Code snippet or statistic
   - Format: Monospace, 28–32pt, white text, dark background
3. **Position:**
   - Drag text box to corner or bottom of screen

Color Correction (if needed)

1. **Select video clip**
2. **Effects → Color → Brightness/Contrast**
3. Adjust if slides appear dark or washed out
4. **Apply**

Audio Mixing

1. **Audio levels check:**
   - Play timeline
   - Monitor Audio Levels (top right of Kdenlive)
   - Target: Narration at –6 to –3 dB
   - Music at –28 dB or lower
2. **Adjust individual track volumes:**
   - Click track name → Volume slider
   - Or use keyframes for fade effects

Preview

1. **Play timeline:** Hit spacebar
2. **Check:**
   - Audio/video sync
   - Timing feels natural (not rushed or slow)
   - Transitions smooth
   - Text readable
   - No awkward silences or gaps

---

7. EXPORT FINAL VIDEO FROM KDENLIVE

1. **File → Export → Export as H.264 MP4**
2. **Settings:**
   - Bitrate: 10000 kbps (10 Mbps, high quality)
   - Audio: AAC 192 kbps
   - Frame rate: 30 fps
   - Resolution: 1920×1080
3. **Filename:** `Palmer_Penguins_Part1_Final.mp4`
4. **Export** (takes 10–20 minutes depending on length)

---

8. GENERATE CAPTIONS WITH OPENAI WHISPER

Installation

Mac/Linux (terminal):

```
pip install openai-whisper
```

Windows: Use WSL or Anaconda

Generate Captions

```
whisper Palmer_Penguins_Part1_Final.mp4 --model large --output_format vtt
```

This creates: `Palmer_Penguins_Part1_Final.vtt`

Review & Edit Captions

1. **Open `Palmer_Penguins_Part1_Final.vtt` in a text editor** (VS Code, Sublime, etc.)

Example format:

```
WEBVTT

00:00:00.000 --> 00:00:08.000
Welcome to Palmer Penguins Part 1: Exploratory Data Analysis...

00:00:08.000 --> 00:00:20.000
I'm [Your Name], and in this series, we're going to explore...
```

2. **Check accuracy:**
   - Verify technical terms (e.g., "Adelie" not "Adley")
   - Fix any obvious speech-to-text errors
   - Adjust timing if needed (timestamps must increment)
3. **Save:**

- File → Save as `Palmer_Penguins_Part1_Final.vtt`

---

9. PROJECT FOLDER STRUCTURE

Organize everything:

```
palmer_penguins_series/
  part1_eda_regression/
    00_analysis_images/         # FROM R ANALYSIS (already generated!)
        penguin-hero-part1.png   # Title slide hero image
        species-distribution.png  # Species bar chart
        island-distribution.png   # Island distribution
        morphometric-distributions.png  # Measurement histograms
        flipper-body-mass-relationship.png  # Main scatter plot
        correlation-matrix.png    # Correlation heatmap
        eda-overview.png          # Combined overview plot
        species-comparison.png    # Body mass boxplot
        simple-regression-model.png  # Regression with CI band
        model-diagnostics.png     # Residual diagnostic plot

    01_planning/
        script_final.md
        lesson_outline.txt
        storyboard.pdf

    02_slides/
        Palmer_Penguins_Part1.odp (LibreOffice original)
        slides_exported/
            slide_01.png
            slide_02.png
            ... (all 18 slides)
        Palmer_Penguins_Part1.pdf

    03_audio/
        narration_raw/
            slide01_narration_take1.wav
            slide01_narration_take2.wav (if multiple takes)
            slide02_narration.wav
            ... (all slides)
        narration_processed/
            slide01_narration.wav (normalized, noise-reduced)
            slide02_narration.wav
            ... (final versions)
        background_music.mp3 (if used)

    04_video_capture/
        slide_capture_raw.mp4 (OBS output)
        obs_settings.txt (configuration reference)

    05_editing/
        Palmer_Penguins_Part1.kdenlive (Kdenlive project)
        renders/
            Palmer_Penguins_Part1_Draft_v1.mp4
```

```
                Palmer_Penguins_Part1_Draft_v2.mp4
                Palmer_Penguins_Part1_Final.mp4
            effects_and_transitions.txt

        06_captions/
            Palmer_Penguins_Part1_Final.vtt (auto-generated)
            Palmer_Penguins_Part1_Final_edited.vtt (corrected)
            transcript.txt (plain text)

        07_deliverables/
            Palmer_Penguins_Part1_Final.mp4
            Palmer_Penguins_Part1_Final_edited.vtt
            Palmer_Penguins_Part1_slides.pdf
            Palmer_Penguins_Part1_transcript.txt
            Palmer_Penguins_Part1_thumbnail.png (1280×720)
            README.md

        08_metadata/
             youtube_description.txt
             youtube_keywords.txt
             creative_commons_license.txt
             code_repository_link.txt

    shared_assets/
        colors.txt (Adelie, Chinstrap, Gentoo hex codes)
        fonts/
            Inter.otf (or system fonts)
            JetBrains_Mono.ttf
        icons/
            penguin_icon.svg
            ... (other icons)
        part_templates/
             slide_template.odp
             script_template.md
```

---

10. THUMBNAIL IMAGE

YouTube thumbnail (1280 × 720 pixels):

Design: In LibreOffice Draw or Figma

```
Background: Gradient (coral #FF6B6B to deep blue #2E86AB)
Center: "Part 1" (large, white, bold)
Bottom: "Palmer Penguins | EDA & Regression"
Corner: Penguin silhouette or icon
```

Export: PNG, 1280×720, save as `Palmer_Penguins_Part1_thumbnail.png`

---

11. YOUTUBE METADATA

SHORT VERSION (3 min)

Title:

Palmer Penguins Part 1: Simple Regression in 3 Minutes

Description:

Can flipper length predict body mass? Let's find out with 333 Antarctic penguins!

Timestamps:
0:00 - The Question
0:10 - Meet the Data (333 penguins, 3 species)
0:30 - Correlation: r = 0.87
0:50 - Simple Regression Model
1:15 - R^2 = 0.762 (76% variance explained)
1:35 - Making Predictions
1:55 - The Problem: Species Clustering
2:20 - Key Takeaways
2:35 - Next: Part 2

Key Results:
• Flipper length explains 76% of body mass variance
• Every 1mm of flipper   50g of body mass
• But species matters-residuals cluster by species!

Resources:
• Full Blog Post: [link]
• Code: [GitHub link]
• Dataset: palmerpenguins R package

Part 2: Adding species → R^2 jumps to 0.86!

#DataScience #Statistics #Regression #Penguins

LONG VERSION (14 min)

Title:

Palmer Penguins Part 1: Exploratory Data Analysis & Simple Regression

Description:

Palmer Penguins Data Analysis Series: Part 1

Learn exploratory data analysis and simple linear regression
using real Antarctic penguin data!

In this video, we:
• Load and explore the Palmer Penguins dataset (333 observations, 3 species)
• Visualize species distributions and morphometric relationships
• Build a simple regression model to predict body mass from flipper length
• Interpret model coefficients and R^2
• Discuss model limitations and next steps

Timestamps:
0:00 - Introduction & Context
1:12 - Why Penguins?
2:20 - Learning Objectives

```
3:00 - Dataset Overview
5:30 - Exploratory Data Analysis
7:45 - Correlation & Key Relationships
9:15 - Introduction to Linear Regression
10:30 - Building the Model
12:00 - Model Results & Interpretation
13:30 - Predictions & Confidence Intervals
14:45 - Model Limitations
15:45 - Key Takeaways
16:30 - Preview of Part 2
```

```
  Resources:
```
- Code Repository: [GitHub link]
- Quarto Blog Post: [Blog link]
- Palmer Penguins Dataset: https://www.kaggle.com/parulpandey/palmer-penguin

```
  This series assumes basic familiarity with R and statistics concepts.
No prior regression experience needed!

Part 2 (Multiple Regression) coming next week!

#DataScience #Statistics #RStatistics #EducationalVideo #OpenScience
```

Keywords

```
palmer penguins, exploratory data analysis, linear regression,
R programming, statistics, data science, educational video,
biostatistics, statistics education, simple regression
```

Captions

Attach the `.vtt` file when uploading

---

12. COMPLETE PRODUCTION TIMELINE

SHORT VERSION (3 min) — Single Day Production

Morning: Slides & Script (1.5 hours) - [ ] Create 8 slides in LibreOffice Impress - [ ] Import the 6 analysis PNG files - [ ] Add titles and minimal text overlays - [ ] Export slides as PNG or PDF

Afternoon: Recording & Editing (2 hours) - [ ] Record 8 narration segments in Audacity (~350 words total) - [ ] Process audio (normalize, noise reduction) - [ ] Import slides + audio to Kdenlive - [ ] Sync timing, add fade transitions - [ ] Export to MP4

Evening: Polish & Upload (1 hour) - [ ] Run Whisper for captions - [ ] Quick caption review - [ ] Create thumbnail - [ ] Upload to YouTube

Total: 4–5 hours for 3-minute video

LONG VERSION (14 min) — Weekly Production

Week 1: Planning & Scripting - [ ] Finalize script (done above) - [ ] Create slide outline (done above) - [ ] Gather any images or data exports needed - **Time: 3–4 hours**

Week 2: Slides & Narration - [ ] Design 18 slides in LibreOffice Impress - [ ] Record narration for all slides (2–3 takes each) - [ ] Process audio (normalize, denoise) in Audacity - [ ] Quality check audio (listen to all) - **Time: 6–8 hours**

Week 3: Video Capture & Editing - [ ] Screen capture slideshow in OBS (or export PNG images) - [ ] Import to Kdenlive - [ ] Assemble timeline (slides + narration audio) - [ ] Add transitions, text overlays - [ ] Color correct if needed - [ ] Export to MP4 - **Time: 5–7 hours**

Week 4: Captions & Finalization - [ ] Run Whisper on final video - [ ] Edit captions for accuracy - [ ] Create thumbnail image - [ ] Write YouTube metadata (description, keywords) - [ ] Do final QA check - [ ] Upload to YouTube - **Time: 3–4 hours**

Total: 17–23 hours for 14-minute video

---

13. FINAL CHECKLIST BEFORE UPLOAD

Content Accuracy: - [ ] All statistics match your R analysis - [ ] Code examples execute correctly (or represent actual output) - [ ] Interpretations are scientifically sound - [ ] Limitations are honestly discussed

Production Quality: - [ ] Audio levels consistent (narration –6 to –3 dB) - [ ] Background music –15 to –18 dB (intro/outro only) - [ ] No background noise or pops - [ ] Video bitrate 10+ Mbps (high quality) - [ ] Resolution 1920×1080 - [ ] Frame rate 30 fps - [ ] No typos in slides or captions

Polish Effects Applied: - [ ] Dissolve transitions between slides (0.3-0.5 sec) - [ ] Ken Burns effect on analysis images (subtle 5% zoom) - [ ] Lower third name overlay on title slide (5-8 sec) - [ ] Fade in at video start (1 sec) - [ ] Fade out at video end (1 sec) - [ ] Music fades: out after intro, in before outro - [ ] No music during educational content (slides 3-17)

Accessibility: - [ ] Captions are accurate (check technical terms) - [ ] Transcript is complete and matches video - [ ] Color palette is colorblind-friendly (you already have this!) - [ ] Text is large and readable (minimum 32pt on slides)

Technical: - [ ] File naming is clear and organized - [ ] Backup copies saved (2 locations) - [ ] YouTube metadata filled in completely - [ ] Links in description are tested and working - [ ] Thumbnail image is 1280×720 PNG

YouTube Setup: - [ ] Video title clear and descriptive - [ ] Description includes timestamps and resources - [ ] Captions (.vtt) uploaded - [ ] Thumbnail image uploaded - [ ] Video categorized as "Education" - [ ] Playlist created: "Palmer Penguins Data Analysis Series" - [ ] License set (CC BY-NC-SA or your choice)

Final Thoughts

This workflow is **100% open-source** and produces **professional-grade** educational content. The same approach used by: - StatQuest (no music during explanations) - 3Blue1Brown (clean visuals, focused narration) - Crash Course (engaging intro/outro)

Key advantages of this approach:

- Slides are the star (clear, readable, minimal)
- Narration drives understanding (conversational, paced)
- Screen captures add depth (code, plots, results)
- Polish effects add professionalism without distraction
- Music only at intro/outro (preserves cognitive load for learning)
- AI avatar option for consistent presenter presence
- Fully reproducible and archivable
- Zero cost (except your time)

Production approach options:

Method	Time	Best For
Audio-only	4-5 hrs	Quick production, content focus
Webcam PiP	5-6 hrs	Personal connection, live feel
AI Avatar hybrid	5-6 hrs	Professional look, no equipment

Good luck with Part 1! Once you complete it, Parts 2–5 follow the same template, making the workflow faster and more efficient.

Questions to Guide Production

- **Does this slide make sense without narration?** (If no, redesign)
- **Is the narration clear at 1.5x speed?** (Yes = good pacing)
- **Can a viewer find the code on GitHub?** (Link it clearly)
- **Would a colleague understand the limitations?** (Explain them)
- **Am I proud to share this?** (If not, do another take)

You've got this!

---

14. CLI-BASED WORKFLOW (ALTERNATIVE TO GUI)

Most of the video production can be done via command line using `ffmpeg` and `sox`. This is faster for batch processing and reproducible builds.

Install CLI Tools

```
# macOS
brew install ffmpeg sox imagemagick poppler

# Linux (Debian/Ubuntu)
sudo apt install ffmpeg sox imagemagick poppler-utils
```

Audio Processing with Sox

```
cd video_production/audio

# Normalize audio to -1dB peak
sox slide01_raw.wav slide01_norm.wav gain -n -1

# Noise reduction (two-step process)
sox slide01_raw.wav -n trim 0 1 noiseprof noise.prof
sox slide01_raw.wav slide01_clean.wav noisered noise.prof 0.21

# Add fade in/out (0.5 sec each)
sox slide01_clean.wav slide01_final.wav fade t 0.5 0 0.5

# Batch process all 18 slides
for i in $(seq -w 1 18); do
  if [ -f "slide${i}_raw.wav" ]; then
    sox slide${i}_raw.wav -n trim 0 1 noiseprof noise.prof 2>/dev/null || true
```

```
    sox slide${i}_raw.wav slide${i}_clean.wav noisered noise.prof 0.21
    sox slide${i}_clean.wav slide${i}_final.wav gain -n -1 fade t 0.3 0 0.3
  fi
done
```

Video Assembly with FFmpeg

```
cd video_production

# Create file list with durations (adjust per slide timing)
cat > slides.txt << 'EOF'
file 'slides_exported/slide_01.png'
duration 8
file 'slides_exported/slide_02.png'
duration 12
file 'slides_exported/slide_03.png'
duration 8
file 'slides_exported/slide_04.png'
duration 10
file 'slides_exported/slide_05.png'
duration 15
file 'slides_exported/slide_06.png'
duration 12
file 'slides_exported/slide_07.png'
duration 8
file 'slides_exported/slide_08.png'
duration 10
file 'slides_exported/slide_09.png'
duration 8
file 'slides_exported/slide_10.png'
duration 10
file 'slides_exported/slide_11.png'
duration 15
file 'slides_exported/slide_12.png'
duration 8
file 'slides_exported/slide_13.png'
duration 12
file 'slides_exported/slide_14.png'
duration 10
file 'slides_exported/slide_15.png'
duration 12
file 'slides_exported/slide_16.png'
duration 10
file 'slides_exported/slide_17.png'
duration 10
file 'slides_exported/slide_18.png'
duration 8
EOF

# Create video from images
ffmpeg -f concat -i slides.txt -vf "scale=1920:1080,format=yuv420p" \
  -c:v libx264 -r 30 slides_only.mp4
```

Concatenate Audio and Combine

```
# Create audio file list
ls -1 audio/slide*_final.wav | sort > audio/audio_list.txt
sed -i '' "s/^/file '/" audio/audio_list.txt
sed -i '' "s/$/'/" audio/audio_list.txt

# Concatenate all audio
ffmpeg -f concat -safe 0 -i audio/audio_list.txt -c:a pcm_s16le audio/narration.wav

# Merge video and audio
ffmpeg -i slides_only.mp4 -i audio/narration.wav \
  -c:v libx264 -c:a aac -b:a 192k \
  -map 0:v:0 -map 1:a:0 \
  video_draft.mp4
```

Add Polish Effects

```
# Get video duration
DURATION=$(ffprobe -v error -show_entries format=duration \
  -of default=noprint_wrappers=1:nokey=1 video_draft.mp4 | cut -d. -f1)
FADE_OUT=$((DURATION - 1))

# Add video and audio fades
ffmpeg -i video_draft.mp4 \
  -vf "fade=t=in:st=0:d=1,fade=t=out:st=${FADE_OUT}:d=1" \
  -af "afade=t=in:st=0:d=0.5,afade=t=out:st=${FADE_OUT}.5:d=0.5" \
  -c:v libx264 -c:a aac \
  video_with_fades.mp4

# Add lower third (first 8 seconds)
ffmpeg -i video_with_fades.mp4 \
  -vf "drawtext=text='Your Name | UCSD Biostatistics':fontsize=28:fontcolor=white:x=80:y=h-80:enable='l
  -c:v libx264 -c:a copy \
  video_with_lower_third.mp4
```

Add Background Music (Intro/Outro Only)

```
# Prepare intro music (20 sec with fade out at end)
ffmpeg -i background_music.mp3 -t 20 \
  -af "afade=t=out:st=19:d=1,volume=-15dB" \
  music_intro.wav

# Prepare outro music (20 sec with fade in at start)
ffmpeg -ss 30 -i background_music.mp3 -t 20 \
  -af "afade=t=in:st=0:d=1,volume=-15dB" \
  music_outro.wav

# Calculate middle section duration (total - intro - outro)
MIDDLE_DURATION=$((DURATION - 40))

# Create silence for middle
ffmpeg -f lavfi -i anullsrc=r=48000:cl=stereo -t $MIDDLE_DURATION silence.wav

# Concatenate music track: intro + silence + outro
```

```
ffmpeg -i music_intro.wav -i silence.wav -i music_outro.wav \
  -filter_complex "[0:a][1:a][2:a]concat=n=3:v=0:a=1[out]" \
  -map "[out]" music_track.wav

# Mix narration with music track
ffmpeg -i audio/narration.wav -i music_track.wav \
  -filter_complex "[0:a][1:a]amix=inputs=2:duration=longest[out]" \
  -map "[out]" final_audio.wav

# Combine final audio with video
ffmpeg -i slides_only.mp4 -i final_audio.wav \
  -c:v libx264 -c:a aac -b:a 192k \
  Palmer_Penguins_Part1_Final.mp4
```

Ken Burns Effect (Per-Slide)

```
# Apply subtle zoom to a single slide (e.g., slide 13 - regression plot)
ffmpeg -loop 1 -i slides_exported/slide_13.png -t 12 \
  -vf "scale=8000:-1,zoompan=z='min(zoom+0.00015,1.05)':x='iw/2-(iw/zoom/2)':y='ih/2-(ih/zoom/2)':d=360
  -c:v libx264 -pix_fmt yuv420p \
  slide13_kenburns.mp4

# Then use this clip instead of static image in final assembly
```

Complete Build Script

Save as build_video.sh:

```
#!/bin/bash
set -e

echo "=== Palmer Penguins Video Builder ==="
echo "Using ffmpeg and sox for CLI-based production"
echo ""

PROJECT_DIR="video_production"
cd "$PROJECT_DIR"

# Step 1: Process audio
echo "Step 1: Processing audio files..."
cd audio
for f in slide*_raw.wav; do
  base="${f%_raw.wav}"
  sox "$f" -n trim 0 1 noiseprof noise.prof 2>/dev/null || true
  sox "$f" "${base}_clean.wav" noisered noise.prof 0.21
  sox "${base}_clean.wav" "${base}_final.wav" gain -n -1 fade t 0.3 0 0.3
  echo "  Processed: $f"
done
cd ..

# Step 2: Concatenate audio
echo "Step 2: Concatenating audio..."
ls -1 audio/slide*_final.wav | sort | sed 's/^/file /' > audio/files.txt
ffmpeg -y -f concat -safe 0 -i audio/files.txt -c:a pcm_s16le audio/narration.wav
```

```bash
# Step 3: Create video from slides
echo "Step 3: Creating video from slides..."
ffmpeg -y -f concat -i slides.txt -vf "scale=1920:1080,format=yuv420p" \
  -c:v libx264 -r 30 slides_only.mp4

# Step 4: Combine video + audio
echo "Step 4: Combining video and audio..."
ffmpeg -y -i slides_only.mp4 -i audio/narration.wav \
  -c:v libx264 -c:a aac -b:a 192k \
  -map 0:v:0 -map 1:a:0 \
  video_draft.mp4

# Step 5: Add fades
echo "Step 5: Adding fade effects..."
DURATION=$(ffprobe -v error -show_entries format=duration \
  -of default=noprint_wrappers=1:nokey=1 video_draft.mp4 | cut -d. -f1)
FADE_OUT=$((DURATION - 1))

ffmpeg -y -i video_draft.mp4 \
  -vf "fade=t=in:st=0:d=1,fade=t=out:st=${FADE_OUT}:d=1" \
  -af "afade=t=in:st=0:d=0.5,afade=t=out:st=${FADE_OUT}.5:d=0.5" \
  -c:v libx264 -c:a aac \
  Palmer_Penguins_Part1_Final.mp4

# Step 6: Generate captions
echo "Step 6: Generating captions..."
whisper Palmer_Penguins_Part1_Final.mp4 --model base --output_format vtt --output_dir .

# Step 7: Create thumbnail
echo "Step 7: Creating thumbnail..."
convert slides_exported/slide_01.png -resize 1280x720! \
  -fill 'rgba(0,0,0,0.5)' -draw 'rectangle 0,550 1280,720' \
  -font Helvetica-Bold -pointsize 64 -fill white \
  -gravity south -annotate +0+90 'Part 1: EDA & Regression' \
  -font Helvetica -pointsize 32 -fill white \
  -gravity south -annotate +0+40 'Palmer Penguins Series' \
  Palmer_Penguins_Part1_thumbnail.png

echo ""
echo "=== Build complete! ==="
echo "Output: Palmer_Penguins_Part1_Final.mp4"
echo "Captions: Palmer_Penguins_Part1_Final.vtt"
echo "Thumbnail: Palmer_Penguins_Part1_thumbnail.png"
```

CLI vs GUI Decision Guide

Task	CLI Best When	GUI Best When
Audio processing	Batch files, consistent settings	Fine-tuning individual clips
Video assembly	Fixed timing, reproducible	Adjusting timing by ear
Transitions	Simple dissolves	Complex transitions
Ken Burns	Same effect on many images	Custom motion per slide
Lower thirds	Same style throughout	Unique per video
Music mixing	Precise dB levels	Adjusting by ear

Task	CLI Best When	GUI Best When
Final review	N/A	Always preview in GUI

Recommended hybrid workflow:

1. Record audio in Audacity (GUI) - better monitoring
2. Process audio with sox (CLI) - consistent quality
3. Assemble video with ffmpeg (CLI) - reproducible
4. Fine-tune timing in Kdenlive (GUI) - visual preview
5. Export final with ffmpeg (CLI) - precise settings