ENVIRONMENTAL CHANGE THROUGH ART: A MEDIUM AND TIME-BASED ANALYSIS

Uttara Parekh

1. Libraries, Functions, Datasets, and Key Variables

1.1 R libraries

Used for:

- **tidyverse** (dplyr, ggplot2, stringr, tidyr, readr, forcats, purrr, tibble, lubridate)
    - Data wrangling, factor handling, string cleaning, and modeling pre-processing.
- **janitor**
    - `clean_names()` to standardize column names (e.g., `"Object ID"` → `object_id`).
- **here**
    - Project-root–based file paths (for reading data reliably).
- **ggplot2**, **ggridges**, **ggfx**
    - Plotting (used mainly in visual part, which we're leaving out here).
- **scales**
    - Better percentage labels and formatting.
- **dplyr**, **forcats**, **stringr**
    - Explicitly loaded (also part of tidyverse) for clarity in pipes, factors, and string matching.

1.2 Custom function

`add_medium_group(df)`

Purpose:
Classify each artwork's raw `medium` text into a broader **medium_group** that reflects material and technological ecology.

How it works:

1. Converts `medium` to lowercase: `medium_lower`.
2. Uses `case_when()` + `str_detect()` to test keywords in `medium_lower`.
3. Assigns a single **medium_group** per row, based on the first matching condition:

    - `"oil"` → **Oil painting**
    - `"watercolor|watercolour|gouache|acrylic"` → **Water-based painting**
    - Direct mark-making on paper, but **not** prints (`pencil|graphite|charcoal|chalk|crayon|pen and ink|ink on paper|wash on paper` **and** NOT `print|etch|engraving|lithograph|woodcut|screenprint|linocut`) → **Drawing / Works on paper**
    - Print-related processes (`print|etch|engraving|lithograph|woodcut|screenprint|linocut|monotype`) → **Printmaking**
    - Photographic processes → **Photography**
    - Digital / video / film / projection / media → **Digital/Media**
    - Sculptural materials (bronze, stone, marble, wood, etc.) → **Sculpture**
    - Everything else → **Other**

Technically: It is a simple wrapper around `mutate()` + `case_when()`; I "feed it a data frame with a `medium` column, get back the same data frame with two new columns: `medium_lower` and `medium_group`".

1.3 Core datasets created in R

Object name	Contents / Role
`met`	Raw Met dataset (448k+ rows) with cleaned column names.
`met_small`	Met subset with only relevant columns (id, title, year, classification, medium).
`met_land`	Met works from ≥1820 that mention environmental keywords and avoid "bad" words.
`met_land_simple`	Met landscape subset with standardized variables + museum label.

Object name	Contents / Role
`met_final`	Met-only subset of **wild** and **urban** works with binary `is_urban`.
`tate_raw`	Raw Tate dataset (~69k rows) with cleaned names.
`tate_small`	Tate subset with id, title, medium, year, classification (constructed).
`tate_land`	Tate works from ≥1820 that mention environmental keywords and avoid "bad" words.
`tate_land_simple`	Tate landscape subset with classification + museum label + depiction type.
`tate_final`	Tate-only subset of **wild** and **urban** works with `is_urban`.
`all_art_full`	Combined Met + Tate landscape universe (wild, urban, mixed, other).
`all_art_model`	Combined dataset restricted to wild + urban, with binary outcome `is_urban`.
`tab_full`	Contingency table of `medium_group × depiction_type`.
`decade_summary`	Decade-level counts and proportion urban across Met + Tate.
`t_test_result`	Welch t-test comparing dates of wild vs urban works.
`chisq_result_full`	Chi-square test of independence between medium and depiction type.
`model1–model4`	Logistic regression models predicting urban vs wild.

1.4 Key variables (columns) created/used

Core identification & description

- `object_id` : Unique ID per artwork (Met: object_id, Tate: id).
- `title` : Artwork title.
- `classification` : Art type / category (Met: from dataset, Tate: from `artist_role`).
- `medium` : Raw text description of material/technique.

Time & phase

- `year` : Production year (Met: `object_begin_date`, Tate: `year`).
- `decade` : `floor(year / 10) * 10` – used for temporal smoothing and decade summaries.
- `phase` : Categorical ecological phase:
  - *Steam & Coal* (1820–1869)
  - *Oil & Empire* (1870–1919)
  - *Motor & Machine Vision* (1920–1959)
  - *Plastic & Suburban Nature* (1960–1989)
  - *Carbon & Crisis* (1990–2010)

Text processing / classification

- `subject_text` : Concatenation of title, object name/classification, and medium.

- `subject_lower` : Lowercase version for case-insensitive keyword matching.
- `env_pattern` : Regex of **environment-related** words (wild + urban).
- `bad_pattern` : Regex of words used to exclude functional / non-landscape objects.
- `is_wild_flag` : `TRUE` if `subject_lower` contains any `wild_words`.
- `is_urban_flag` : `TRUE` if `subject_lower` contains any `urban_words`.
- `depiction_type` : Final 4-level classification: **wild**, **urban**, **mixed**, **other**.

Material ecology

- `medium_lower` : Lowercased `medium` used inside `add_medium_group()`.
- `medium_group` : Broader material category (Oil painting, Printmaking, etc.).

Institution / modeling

- `museum` : Factor with levels `"Met"` and `"Tate"`.
- `is_urban` : Binary outcome (1 = urban, 0 = wild) used in logistic regression.
- `prop_urban` : Proportion of artworks that are urban in each decade (for `decade_summary`).

---

2. Data Filtering and Construction of the Analysis Universe

2.1 Environmental keyword filtering

I built a consistent **text-based environmental filter** for both Met and Tate:

1. **Positive filters**
    - `environment_words` = `landscape_words` + `urban_words`
    - Capture rivers, mountains, forests, fields, cities, bridges, roads, factories, ports, etc.
2. **Negative filters (`bad_words`)**
    - Exclude objects like coins, medals, bowls, vases, furniture, masks, armor, jewelry, etc.
    - These are often functional or decorative; even if they mention “river” in the description, they’re not landscape images.
3. **Rules applied** (Met as example; same for Tate):

```
met_land <- met_small %>%
  filter(!is.na(year)) %>%
  filter(year >= 1820) %>%
  filter(str_detect(subject_lower, env_pattern)) %>%
  filter(!str_detect(subject_lower, bad_pattern))
```

This ensures:

- Only dated artworks from 1820 onwards,
- With at least one environmental keyword,
- And no "bad" object-type keywords.

2.2 Wild, urban, mixed, other

I then **tag each artwork**:

```
is_wild_flag  = str_detect(subject_lower, paste(wild_words,  collapse = "|"))
is_urban_flag = str_detect(subject_lower, paste(urban_words, collapse = "|"))

depiction_type = case_when(
  is_wild_flag & !is_urban_flag ~ "wild",
  !is_wild_flag & is_urban_flag ~ "urban",
  is_wild_flag & is_urban_flag  ~ "mixed",
  TRUE                          ~ "other"
)
```

- **wild** = only wild terms.
- **urban** = only urban terms.
- **mixed** = both wild & urban terms → transitional or hybrid environments.
- **other** = ambiguous, minimal, or non-environmental language.

Resulting distribution (combined Met + Tate):

Depiction type	*Count*
Wild	8,821
Urban	9,149
Mixed	4,506
Other	4,547

This becomes the **full analysis universe**: `all_art_full`.

2.3 Medium grouping and ecological phases

After classification, I:

« Apply `add_medium_group()` to both `met_land_simple` and `tate_land_simple`, creating:

- Oil painting
- Water-based painting
- Drawing / Works on paper
- Printmaking
- Photography
- Digital/Media
- Sculpture
- Other

« Create **decade** and **phase**:

```
decade = floor(year / 10) * 10
phase = case_when(
  year >= 1820 & year < 1870 ~ "Steam & Coal",
  year >= 1870 & year < 1920 ~ "Oil & Empire",
  year >= 1920 & year < 1960 ~ "Motor & Machine Vision",
  year >= 1960 & year < 1990 ~ "Plastic & Suburban Nature",
  year >= 1990               ~ "Carbon & Crisis",
  TRUE ~ NA_character_
)
```

These phases are historically grounded in energy regimes and environmental history: coal and steam, petroleum and empire, motorization and mechanical vision, plastics and postwar suburbs, late-century carbon crisis.

2.4 Combined datasets

- `all_art_full` = `bind_rows(met_land_simple, tate_land_simple)` with factors for `museum`, `medium_group`, `phase`, `depiction_type`.
- `all_art_model` = `all_art_full` restricted to **wild + urban** and with `is_urban` binary:

```
all_art_model <- all_art_full %>%
  filter(depiction_type %in% c("wild", "urban")) %>%
  mutate(
    is_urban = if_else(depiction_type == "urban", 1L, 0L),
    depiction_type = droplevels(depiction_type)
  )
```

This is the core dataset for all inferential tests.

---

3. Descriptive Patterns

3.1 Range of years by museum

```
all_art_full %>%
  filter(!is.na(year)) %>%
  group_by(museum) %>%
  summarise(min_year = min(year), max_year = max(year))
```

Museum	*Min year*	*Max year*
Met	1820	2016
Tate	1820	2010

Interpretation:
Both collections cover the entire period of interest (1820 onwards), with the Met extending slightly further into the 21st century. This matters because late peaks in urban imagery may be slightly more pronounced at the Met.

3.2 Medium × depiction type (row-wise proportions)

Computing a contingency table:

```
tab_full <- table(all_art_full$medium_group, all_art_full$depiction_type)
prop.table(tab_full, margin = 1)
```

Key proportions (within each medium):

Medium Group	*Wild*	*Urban*	*Mixed*	*Other*
Digital/Media	0.105	0.658	0.177	0.060
Drawing / Works on paper	0.374	0.191	0.253	0.181
Oil painting	0.387	0.222	0.087	0.304
Photography	0.657	0.215	0.091	0.036
Printmaking	0.297	0.412	0.138	0.153
Sculpture	0.202	0.585	0.074	0.138
Water-based painting	0.428	0.175	0.138	0.260

Interpretation:

- Digital/Media and Sculpture lean strongly **urban**.
- Photography, Drawing, and Water-based painting lean strongly **wild** or **mixed**.
- Printmaking sits somewhere in between, with a noticeable urban skew.

3.3 Decade summary of urban share

```
decade_summary <- all_art_model %>%
  group_by(decade) %>%
  summarise(
    n        = n(),
    n_urban  = sum(is_urban),
    prop_urban = n_urban / n
  )
```

Selected highlights:

- 1820s: prop_urban ≈ 0.31
- 1850s: prop_urban ≈ 0.56
- 1920s: prop_urban ≈ 0.71
- 1930s: prop_urban ≈ 0.74
- 1940s: prop_urban ≈ 0.77
- 1970s: prop_urban ≈ 0.74

Interpretation:
Across the 19th and 20th century, the share of urban landscapes steadily increases, with particularly high urban fractions in the interwar and mid–20th-century decades.

---

4. Inferential Analyses

4.1 T-test: Are urban works newer than wild works?

Model: `t.test(year ~ depiction_type, data = all_art_model)`

Group	Mean year
Wild	1880.4
Urban	1907.8

- **Difference**: ~27 years
- **t** = −36.57
- **df** ≈ 17,917
- **p** < 2.2e−16
- **95% CI for difference (urban − wild)**: [25.89, 28.82] years

Interpretation:
Urban landscapes appear, on average, almost **three decades later** in the historical record than wild landscapes, and this difference is statistically extremely strong. This aligns with the idea that industrialization and urban growth are later developments in both environmental history and museum collections.

4.2 Chi-square: Are medium and depiction type independent?

Model: `chisq.test(tab_full)`

- $\chi^2(21) = 3713.5$
- $p < 2.2e{-}16$

Interpretation:
There is a **very strong association** between artistic medium and how the environment is depicted (wild, urban, mixed, other). Medium is not just a technical choice—it strongly structures whether landscapes appear as wilderness, city, or something in between.

4.3 Logistic regression models (urban vs wild)

Outcome: `is_urban` (1 = urban, 0 = wild)
Link: logit (log-odds)

Model 1: Time-only model

```
model1 <- glm(is_urban ~ year, family = "binomial", data = all_art_model)
```

Term	*Estimate (β)*	*SE*	*z*	*p-value*	*Odds Ratio (OR)*
Intercept	−20.04	0.581	−34.47	< 2e−16	1.99e−09
Year	+0.0106	0.00031	+34.53	< 2e−16	**1.0107**

Interpretation:

- Each additional year increases the odds of an artwork being urban by ~1.07%.
- Over 100 years, the odds of urban depiction roughly increase by about **49%** (OR ≈ 1.49).
- Time alone explains a substantial portion of the shift from wild to urban imagery.

Model 2: Medium-only model

```
model2 <- glm(is_urban ~ medium_group, family = "binomial", data =
all_art_model)
```

Reference category: **Digital/Media**

Medium Group (vs Digital/Media)	*β*	*SE*	*z*	*p-value*	*OR (urban odds vs Digital/Media)*
Intercept (Digital/Media)	1.834	0.063	28.91	< 2e−16	**6.26**
Drawing / Works on paper	−2.505	0.071	−35.09	< 2e−16	**0.082**
Oil painting	−2.387	0.097	−24.54	< 2e−16	**0.092**
Photography	−2.949	0.116	−25.50	< 2e−16	**0.052**
Printmaking	−1.508	0.068	−22.31	< 2e−16	**0.221**
Sculpture	−0.771	0.166	−4.64	3.5e−06	**0.462**
Water-based painting	−2.730	0.086	−31.86	< 2e−16	**0.065**

Interpretation:

- Digital/Media is the most **urban-leaning** baseline.
- All other media have **much lower odds** of urban representation compared with Digital/Media.
- Photography and Water-based painting are especially wild-leaning (OR ~0.05–0.07 vs Digital/Media).

Model 3: Year + medium + ecological phase

```
model3 <- glm(is_urban ~ year + medium_group + phase, family = "binomial",
data = all_art_model)
```

Key coefficients (odds ratios):

Predictor	*β*	*OR*	*Interpretation*
Year	0.0041	1.0041	Urban odds slowly increase over time (controlling for medium and phase).
Drawing / Works on paper	−1.904	0.149	Much less urban than Digital/Media.
Oil painting	−2.074	0.126	Less urban than Digital/Media.
Photography	−2.382	0.092	Strong wild orientation vs Digital/Media.
Printmaking	−1.209	0.298	Mixed but still less urban than Digital/Media.
Water-based painting	−2.224	0.108	Strong wild orientation.
Phase: Oil & Empire	−0.236	0.79	Lower odds of urban vs Steam & Coal.

Phase: Motor & Machine Vision	0.199	1.22	Slightly higher urban odds (ns at 0.05).
Phase: Plastic & Suburban Nature	0.347	1.41	Higher urban odds (borderline at 0.086).
Phase: Carbon & Crisis	−0.600	0.55	Lower urban odds compared with Steam & Coal.

Interpretation:

- Even after controlling for medium and phase, urban depictions increase with time.
- Certain phases (especially Plastic & Suburban Nature) are more urban; others (Oil & Empire, Carbon & Crisis) are less so.
- Medium remains a strong structural predictor.

Model 4: Full model (year + medium + phase + museum)

```
model4 <- glm(is_urban ~ year + medium_group + phase + museum, family =
"binomial", data = all_art_model)
```

Key odds ratios:

Predictor	***β***	***OR***	***Interpretation***
Year	0.00096	1.001	Time effect weak once medium, phase, and museum are added.
Drawing / Works on paper	−1.602	0.201	Less urban vs Digital/Media.
Oil painting	−1.883	0.152	Less urban vs Digital/Media.
Photography	−2.323	0.098	Strong wild orientation.
Printmaking	−1.193	0.303	Mixed but urban-leaning compared with wild media.
Water-based painting	−2.038	0.130	Wild-leaning.
Phase: Oil & Empire	−0.258	0.77	Less urban odds than Steam & Coal.
Phase: Motor & Machine Vision	0.350	1.42	Higher urban odds.
Phase: Plastic & Suburban Nature	0.713	2.04	≈2× higher odds of urban depiction.
Phase: Carbon & Crisis	−0.094	0.91	No strong difference vs Steam & Coal.
Museum: Tate (vs Met)	−0.549	**0.58**	Tate has ~42% lower odds of urban depiction than the Met.

Interpretation:

- The **best-fitting model** (lowest AIC = 22098) shows that **medium, ecological phase, and institution** jointly structure whether landscapes are urban or wild.
- Once those are accounted for, the pure effect of year becomes small, meaning "time" matters mainly through changes in **technology, environmental phases, and collecting histories**.

---

5. Interpretation in Context

5.1 Time and environmental imagination

- Urban images appear about **27 years later** than wild images on average.
- Urban proportions climb sharply from late 19th to mid-20th century (peaking in 1920–1970 decades).
- This trajectory aligns with industrialization, railway and roadway expansion, and the growth of metropolitan centers, which gradually transform both real landscapes and their visual representation.

5.2 Medium as ecological technology

Medium isn't just a neutral format; it acts as a **technology of seeing**:

- **Digital/Media & Sculpture**: strongly urban, tied to contemporary infrastructures, architecture, and public art contexts.
- **Photography**: surprisingly wild-oriented here—possibly reflecting documentary and conservationist uses of photography for wilderness and environmental subjects.
- **Water-based painting & Drawings**: closely associated with wild and mixed landscapes—sketching, open-air studies, and watercolor fieldwork retain strong ties to rivers, hills, and forests.
- **Printmaking** sits between worlds, servicing both pastoral and urban scenes.

5.3 Ecological phases and museums

- **Plastic & Suburban Nature (1960–1989)**: ~2× higher odds of urban imagery (OR ≈ 2.04 vs Steam & Coal). This phase overlaps with highways, mass car culture, large-scale urban growth, and environmental crises entering public consciousness.
- **Motor & Machine Vision (1920–1959)**: elevated urban odds, consistent with aerial views, industrial landscapes, and mechanical reproduction of city scenes.
- **Oil & Empire** and **Carbon & Crisis** exhibit more complex patterns, with lower odds of urban depiction once other variables are accounted for—likely reflecting a mix of imperial landscapes, colonial pastoral views, late environmental critique, and renewed interest in wilderness imagery.

- **Institutional differences**: Even holding year, medium, and phase constant, **Tate** exhibits **lower odds of urban depiction** than the Met (OR ≈ 0.58), suggesting distinct curatorial priorities and collection histories.

---

6. Limitations

« **Keyword-based classification**

- Relies on titles/medium text; images without explicit landscape words are missed, and some ambiguous phrases may be misclassified.

« **Language and bias**

- Word lists reflect Anglophone descriptors of environment; non-English or metaphorical titles may be under-detected.

« **Collection bias**

- Met and Tate collections are not random samples of global art; they reflect Western, institutional priorities shaped by acquisition histories.

« **Temporal imbalance**

- Earlier centuries and certain media are more heavily represented than others. Later decades (especially 2000s) have fewer works, which may exaggerate apparent peaks or dips.

---

7. Scope and Future Work

A natural next step would be to **link environmental depictions to artist geographies and acquisition histories**—for example, comparing wild vs urban imagery by artist birthplace, region of production, or by tracing when specific works entered the museum. Extending the method to other museums or to non-Western collections could test whether the same patterns hold globally.

---

8. Conclusion

Museum landscapes shift dramatically from the 19th to the 20th century, moving from predominantly wild environments to increasingly urban ones. Urban works appear, on average,

about **27 years later** than wild works (1880.4 vs 1907.8), and the share of urban imagery peaks around the mid–20th century. This temporal shift is not random: it tracks industrial development, infrastructure expansion, and changing environmental awareness.

Across media, the project shows that **material choices embed ecological choices**. Digital/Media and sculptural works lean strongly toward urban subjects, while Drawing and Water-based painting remain more closely tied to wild or mixed environments. Photography and printmaking occupy hybrid roles, sometimes documenting wilderness, sometimes foregrounding the built environment.

The best-fitting regression model (AIC ≈ 22098) demonstrates that the probability of an artwork depicting urban rather than wild nature is jointly shaped by **medium, ecological phase, and institution**, with time playing a secondary role once these factors are considered. Works in the Plastic & Suburban Nature phase have about **twice the odds** of urban depiction compared to the Steam & Coal era, and Tate shows about **42% lower odds** of urban imagery than the Met after controlling for other variables.

Together, these findings suggest that large museum collections quietly mirror real-world environmental change: as societies industrialize, urbanize, and confront ecological crisis, their visual archives transition from wilderness and pastoral vistas toward cities, infrastructure, and contested environments—mediated by the technologies and materials through which art is made.