```python
# Simple PIAAC SPSS to R Converter (Robust Version)
# ================================================

import pandas as pd
import numpy as np
import pyreadstat
import os

print("Simple PIAAC SPSS to R Converter")
print("=" * 40)

# Step 1: Check if file exists and try to load
print("1. Checking for SPSS file...")

filename = 'prgusap1_puf.sav'
if not os.path.exists(filename):
    print(f"❌ File '{filename}' not found!")
    print("Available .sav files in current directory:")
    sav_files = [f for f in os.listdir('.') if f.endswith('.sav')]
    if sav_files:
        for f in sav_files:
            print(f"   - {f}")
        print(f"\nIf your file has a different name, update the 'filename' variable above.")
    else:
        print("   No .sav files found")
    print("Stopping execution.")
else:
    print(f"✓ Found file: {filename}")

    # Step 2: Load the file
    print("\n2. Loading SPSS file...")
    try:
        # Simple load first - just get the data
        df, meta = pyreadstat.read_sav(filename)
        print(f"✓ Successfully loaded: {df.shape[0]:,} rows × {df.shape[1]:,} columns")

        # Step 3: Quick export to CSV (most important)
        print("\n3. Exporting to CSV...")
        df.to_csv('piaac_full_dataset.csv', index=False)
        file_size = os.path.getsize('piaac_full_dataset.csv') / 1024**2
        print(f"✓ Exported: piaac_full_dataset.csv ({file_size:.1f} MB)")

        # Step 4: Create a subset with key research variables
        print("\n4. Creating research subset...")

        # Define key variables for your research
        key_vars = []

        # Essential variables
        essential = ['SEQID', 'SPFWT0', 'GENDER_R', 'AGEG10LFS_T', 'EDCAT8', 'PARED']
        for var in essential:
            if var in df.columns:
                key_vars.append(var)

        # Plausible values
```

```
        for domain in ['LIT', 'NUM', 'PSL']:
            for i in range(1, 11):
                var = f'PV{domain}{i}'
                if var in df.columns:
                    key_vars.append(var)

        # First 10 replicate weights
        for i in range(1, 11):
            var = f'SPFWT{i}'
            if var in df.columns:
                key_vars.append(var)

        # Create subset
        if key_vars:
            subset_df = df[key_vars].copy()
            subset_df.to_csv('piaac_research_subset.csv', index=False)
            print(f"✓ Research subset: {len(key_vars)} variables saved")

        # Step 5: Create simple R script
        print("\n5. Creating R import script...")

        r_script = f'''# Load PIAAC Data in R
# ==================

# Load required library
library(readr)

# Load the full dataset
cat("Loading PIAAC data...\\n")
piaac_data <- read_csv("piaac_full_dataset.csv")

# Or load just the research subset (faster)
# piaac_data <- read_csv("piaac_research_subset.csv")

# Basic information
cat("Dataset dimensions:", nrow(piaac_data), "rows ×", ncol(piaac_data), "columns\\n")

# Check for key variables
key_vars <- c("SEQID", "SPFWT0", "EDCAT8", "PARED", "PVLIT1", "PVNUM1", "PVPSL1")
available_vars <- key_vars[key_vars %in% names(piaac_data)]
cat("Key variables found:", length(available_vars), "of", length(key_vars), "\\n")
cat("Available:", paste(available_vars, collapse=", "), "\\n")

# For survey analysis, install and load survey package
# install.packages("survey")
# library(survey)
#
# # Create survey design object
# piaac_design <- svydesign(ids = ~1, weights = ~SPFWT0, data = piaac_data)

cat("\\nData loaded successfully! Ready for analysis.\\n")
'''

        with open('load_piaac_in_r.R', 'w') as f:
            f.write(r_script)
        print("✓ Created: load_piaac_in_r.R")
```

```
    # Step 6: Create variable list
    print("\n6. Creating variable list...")

    var_info = []
    for col in df.columns:
        var_info.append({
            'variable': col,
            'type': str(df[col].dtype),
            'missing_count': df[col].isnull().sum(),
            'unique_values': df[col].nunique() if df[col].nunique() < 50 else '>50'
        })

    var_df = pd.DataFrame(var_info)
    var_df.to_csv('piaac_variables.csv', index=False)
    print(f"✓ Variable list: {len(var_df)} variables documented")

    # Summary
    print(f"\n" + "="*40)
    print("SUCCESS! Files created:")
    print("✓ piaac_full_dataset.csv - Complete dataset")
    if key_vars:
        print("✓ piaac_research_subset.csv - Key research variables")
    print("✓ piaac_variables.csv - Variable information")
    print("✓ load_piaac_in_r.R - R import script")
    print(f"\nTo use in R:")
    print(f"1. Open R/RStudio")
    print(f"2. Set working directory to this folder")
    print(f"3. Run: source('load_piaac_in_r.R')")
    print(f"4. Your data will be in 'piaac_data'")

except Exception as e:
    print(f"❌ Error loading file: {e}")
    print("This might be due to:")
    print("- File corruption")
    print("- Insufficient memory")
    print("- Incompatible SPSS version")
    print("\nTry loading the file directly in R instead:")
    print("library(haven)")
    print("piaac_data <- read_sav('prgusap1_puf.sav')")
```